

# ADP37B7IAIVV BEST

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## SUMMARIES

and is derived by analysis of the total score distribution.

OM protein - protein search, using SW model

Run on: January 6, 2004, 09:36:09 ; Search time 47 Seconds

(without alignments)  
1644.677 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407  
Sequence: 1 MHINSAQQPGVAMSEFRT.....EDGGISAFNSRSTPQLRL 487

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*\*

1: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1990.DAT:\*  
11: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1991.DAT:\*  
12: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1992.DAT:\*  
13: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1993.DAT:\*  
14: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1994.DAT:\*  
15: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1995.DAT:\*  
16: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1996.DAT:\*  
17: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1997.DAT:\*  
18: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1998.DAT:\*  
19: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA1999.DAT:\*  
20: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA2000.DAT:\*  
21: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA2001.DAT:\*  
22: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA2002.DAT:\*  
23: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA2003.DAT:\*  
24: /SIDS1/gcgcata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

Result No.	Score	Query Match	Length	DB ID	Description
1	2407	100.0	487	AAE12603	Pseudomonas syring
2	1869.5	77.7	486	AAE12573	Pseudomonas syring
3	1869.5	77.7	486	AAE67677	Amino acid sequenc
4	147	6.1	1463	AAE20110	Lactobacillus rham
5	146.5	6.1	2122	ABU08784	Moraxella catarrha
6	146.5	6.1	2123	AAE00701	Moraxella catarrha
7	143	5.9	10431	ABU54861	Human C125 emlno
8	140	5.8	2285	AAU98149	Bacillus subtilis
9	138	5.7	1229	ABR71736	Drosophila melanog
10	138	5.7	1277	AAE76536	Corynebacterium gl
11	138	5.7	2012	AAE76534	Corynebacterium gl
12	138	5.7	2993	AAE92485	C glutamicum prote
13	136	5.7	1026	AAE48993	rfaA S-lyase prote
14	132	5.5	1026	AAW37490	Caulobacter cresce
15	132	5.5	1026	AAV44757	Caulobacter cresce
16	132	5.5	6077	ABP95996	White spot syndrom
17	130.5	5.4	2368	ABJ19059	Pathogen specific
18	130.5	5.4	2368	AAU34139	Staphylococcus aur
19	130.5	5.4	2368	AAU36796	Staphylococcus aur
20	130	5.4	1026	AAE94014	Caulobacter S-layer
21	130	5.4	1026	AAE34374	Caulobacter cresce
22	129.5	5.4	2076	AAU34319	Staphylococcus aur
23	129.5	5.4	2186	AAU37320	Staphylococcus aur
24	129	5.4	779	ABP65343	Staphylococcus aur
25	129	5.4	2344	AAU37120	Staphylococcus aur
26	129	5.4	6025	AAE84939	Shrimp white spot
27	127.5	5.3	2387	AAU01183	Rat glutamate tran
28	127.5	5.3	2478	AAU34320	Staphylococcus aur
29	127.5	5.3	2478	AAU37374	Staphylococcus aur
30	127.5	5.3	2478	AAU37374	Staphylococcus aur
31	126	5.2	727	ABE25530	Novel human diagno
32	125.5	5.2	623	AAU12038	Clostridium diffic
33	125.5	5.2	1252	AAE80530	B. spheraeaus SLP
34	125	5.2	456	ABE64582	Drosophila melanog
35	125	5.2	793	AAU27571	Neisseria meningit
36	125	5.2	793	AAE10016	N. meningitidis st
37	125	5.2	2086	AAU34143	Staphylococcus aur
38	125	5.2	5795	AAU37017	Staphylococcus aur
39	124	5.2	628	AAE66009	F. necrophorum tru
40	124	5.2	630	ABJ26539	Aspergillus fumiga
41	124	5.2	3241	AAE66005	F. necrophorum leu
42	124	5.2	3647	AAE05041	Filamentous haemag
43	123.5	5.1	2434	AAU34339	Staphylococcus aur
44	123.5	5.1	6281	AAU37403	Staphylococcus aur
45	123.5	5.1	10498	ABJ19119	Pathogen specific

## ALIGNMENTS

RESULT 1  
AAE12603  
ID AAE12603 standard; Protein; 487 AA.  
XX  
AC AAE12603;  
XX  
DT 03-JAN-2002 (first entry)  
XX  
DE Pseudomonas syringae pv. tomato strain DC3000 HopPtoA2 protein.  
XX  
KW Conserved Effector Loci; CEL; cytosolic; antibacterial; gene therapy;  
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;  
KW eukaryotic cell death; cancer.  
XX  
OS Pseudomonas syringae.  
XX  
PN W0200175066-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 03-APR-2001; 2001WO-US10698.  
XX  
PR 03-APR-2000; 2000US-19416P.  
PR 11-AUG-2000; 2000US-224604P.  
PR 17-NOV-2000; 2000US-249348P.  
XX  
PA (CORR.) CORNELL RES FOUND INC.  
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.  
PA (UYNE-) UNIV NEBRASKA.  
PI Collier A, Alfano JR, Charkowski AO;  
XX  
DR WPI1; 2001-639361/73.  
DR N-PSDB; AAD20438.  
XX  
PT New nucleic acid molecules encoding proteins or polypeptides of  
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci  
PT genomic sequences, for imparting disease resistance to plants -  
XX  
PS Claim 8; Page 65-66; 217pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule comprising a  
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas  
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)  
CC genomic sequences. CEL and EEL DNA are useful for imparting disease  
CC resistance to a plant, by transforming a plant cell with the nucleic acid  
CC and regenerating a transgenic plant from the transformed plant cell,  
CC where the transgenic plant expresses a heterologous DNA molecule under  
CC conditions effective to impart disease resistance, or by treating a plant  
CC with an isolated protein or polypeptide, by applying the protein or  
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria  
CC which secretes the protein or polypeptide, under conditions effective to  
CC impart disease resistance to the treated plant. CEL and EEL proteins  
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic  
CC Pseudomonas protein into a eukaryotic cell under conditions effective to  
CC cause cell death. CEL and EEL proteins are also useful for treating a  
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into

CC cancer cells of a patient under conditions effective to cause death of  
CC cancer cells, and thus treating the cancerous condition. The method  
CC further involves administering a targeted DNA delivery system  
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,  
CC to the patient, where the targeted DNA delivery system delivers the  
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein  
CC is expressed in the cancer cells. The present sequence is  
CC Pseudomonas syringae pv. syringae HopPtoA2 homolog protein.  
XX  
SQ Sequence 487 AA;  
Query Match 100.0%; Score 2407; DB 22; Length 487;  
Best Local Similarity 100.0%; Pred. No. 6.3e-192;  
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MHINSAQQPGVMESEFRTASDASLASSVSRYSTTSQRDLQAITDYLMKHVFAAHFSS 60  
DB 1 MHINSAQQPGVMESEFRTASDASLASSVSRYSTTSQRDLQAITDYLMKHVFAAHFSS 60  
QY 61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKKEPDRLATTA5SAF 120  
DB 61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKKEPDRLATTA5SAF 120  
QY 121 ENTFFAASVLYQWMPAINKGIMLTPLEPLISGALSQANQVGTAKMDRAGDLHY 180  
DB 121 ENTFFAASVLYQWMPAINKGIMLTPLEPLISGALSQANQVGTAKMDRAGDLHY 180  
QY 181 LSTSPDKLHDMAVSVKSHSPALGROVDMGIAVQTF5ALNVVTVLAAPLASRSYOGA 240  
DB 181 LSTSPDKLHDMAVSVKSHSPALGROVDMGIAVQTF5ALNVVTVLAAPLASRSYOGA 240  
QY 241 VDEGVSTAGGLVANAQFGDRMLSVQSRDQLRGAFVLYGKKDEPKAALSSETDWLDAYKA 300  
DB 241 VDEGVSTAGGLVANAQFGDRMLSVQSRDQLRGAFVLYGKKDEPKAALSSETDWLDAYKA 300  
QY 301 IKSASYSGAALNAGKRMAGLEPLDYATDGLKAVRSLSATSILTRIGLALAGYAGVSKLOK 360  
DB 301 IKSASYSGAALNAGKRMAGLEPLDYATDGLKAVRSLSATSILTRIGLALAGYAGVSKLOK 360  
QY 361 MATKNIITDSATKAASQSLNLSVGVGFAGMTTAGLATPDAVKRAESFIODKVKSTASST 420  
DB 361 MATKNIITDSATKAASQSLNLSVGVGFAGMTTAGLATPDAVKRAESFIODKVKSTASST 420  
QY 421 TSYVADQTVKLATTVKIDSGEALISSTGASLSTVNNLRRHSAPKADI EEGGISAFFSRSET 480  
DB 421 TSYVADQTVKLATTVKIDSGEALISSTGASLSTVNNLRRHSAPKADI EEGGISAFFSRSET 480  
QY 481 PFOLRRL 487  
DB 481 PFOLRRL 487  
RESULT 2  
AAE12573  
ID AAE12573 standard; Protein; 486 AA.  
XX  
AC AAE12573;

XX 03-JAN-2002 (first entry)  
 DT Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 encoded protein.  
 XX  
 DE  
 XX Conserved Effector Loci; CEL; cytosolic; antibacterial; gene therapy;  
 KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;  
 KW eukaryotic cell death; cancer.  
 XX  
 OS Pseudomonas syringae.  
 XX  
 PN W0200175066-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 03-APR-2001; 2001WO-US10698.  
 XX  
 PR 03-APR-2000; 2000US-19416P.  
 PR 11-AUG-2000; 2000US-224604P.  
 PR 17-NOV-2000; 2000US-249548P.  
 XX  
 PA (CORR.) CORNELL RES FOUND INC.  
 PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.  
 PA (UYNE-) UNIV NEBRASKA.  
 XX  
 PI Collier A, Alfano JR, Charkowski AO;  
 XX  
 XX WPI; 2001-639361/73.  
 DR N-PSDB; AAD20408.  
 XX  
 XX Claim 8; Page 21-23; 217pp; English.  
 PS  
 PT New nucleic acid molecules encoding proteins or polypeptides of  
 PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci  
 PT genomic sequences, for imparting disease resistance to plants ~  
 XX  
 CC The invention relates to an isolated nucleic acid molecule comprising a  
 CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas  
 CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)  
 CC genomic sequences. CEL and EEL DNA are useful for imparting disease  
 CC resistance to a plant, by transforming a plant cell with the nucleic acid  
 CC and regenerating a transgenic plant from the transformed plant cell,  
 CC where the transgenic plant expresses a heterologous DNA molecule under  
 CC conditions effective to impart disease resistance, or by treating a plant  
 CC with an isolated protein or polypeptide, by applying the protein or  
 CC polypeptide in an isolated form or by applying a non-pathogenic bacteria  
 CC which secretes the protein or polypeptide, under conditions effective to  
 CC impart disease resistance to the treated plant. CEL and EEL proteins  
 CC are useful for causing eukaryotic cell death, by introducing a cytotoxic  
 CC Pseudomonas protein into a eukaryotic cell under conditions effective to  
 CC cause cell death. CEL and EEL proteins are also useful for treating a  
 CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into  
 CC cancer cells, and thus treating the cancerous condition. The method  
 CC further involves administering a targeted DNA delivery system  
 CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,  
 CC to the patient, where the targeted DNA delivery system delivers the

CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein  
 CC is expressed in the cancer cells. The present sequence is  
 CC Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 protein.  
 XX  
 SQ Sequence 486 AA;  
 Query Match 77.7%; Score 1869.5; DB 22; Length 486;  
 Best Local Similarity 78.4%; Pred. No. 3.8e-147;  
 Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;  
 QY 1 MHNOSAOQPPGVAFNESEFRITSDSLASSVSRYSTTSRCDLOAITDYLKHHVFAARFS 60  
 DB 1 MHNNRVQQPPVATDSFRITSDSLASSVSRYSTTSRCDLOAITDYLKHHVFAARFS 60  
 QY 61 VIGSPDERDALAHNEQIDALVETRANRLYSEGETPATIETFAKAEKTRDLATTASSAF 120  
 DB 61 PADSDAQGANVDHNAQITALETRASRLHFEGETPATIADTFPAKAEKTRDLATTSSAL 120  
 QY 121 ENTFFAAASVLQYMPALINKGDMLATPLKPLFLISGALSQAMQVGTAKMDRANGDLHY 180  
 DB 121 RATPFMAASLLQYMPALINKGDMLATPLKPLFLISGALSQAMQVGTAKMDRANGDLHY 180  
 QY 181 LSTSPDKLHDMASVSKHSPALGROVDMGIAVQTFESALNVVSTVLAALASRPSYOGA 240  
 DB 181 LSAFSPDRLHDMASVSKHSPALGROVDMGIAVQTFESALNVVSTVLAALASRPAVOGA 240  
 QY 241 VDEGVSTAGLVANNAFGDRMLSVQSRDQLRGAFLVAKDKERPAALSEETDMDAYKA 300  
 DB 241 VDLGVSNAGLVANNAFGDRMLSVQSRDQLRGAFLVAKDKERPAALSEETDMDAYKA 300  
 QY 301 IKASYSQALNNGKRNAGLEPLDVAITDGLKAVBSLVASATSLRIGLALAGVAGVSKLQK 360  
 DB 301 IKASYSQALNNGKRNAGLEPLDVAITDGLKAVBSLVASATSLRIGLALAGVAGVSKLQK 360  
 QY 361 MATKNTDTSATKAASQSLNLYSGVGFAGTGTGLATDPAVKRAESFIODVVKSTAST 420  
 DB 361 MATKNTDTPATKAASQSLNLYSGVGFAGTGTGLATDPAVKRAESFIODVVKSTAST 420  
 QY 421 TSYVADQTVKLAKTVKIDSGEALISSTGASLSTVNNLRHSAPENDIEEGISAFSRSET 480  
 DB 421 TGVVADQTVKLAKTVKIDSGEALITHTGASLNTVNNLRGPAREADIEEGTAA-SPSEI 479  
 QY 481 PQLAR 486  
 DB 480 PFRPFR 485  
 RESULT 3  
 AAB67677  
 ID AAB67677 standard; Protein; 486 AA.  
 XX  
 AC AAB67677;  
 XX  
 DT 11-JUN-2001 (first entry)  
 XX  
 XX Amino acid sequence of a HopProA protein of Pseudomonas syringae.  
 DB  
 XX

KW HIV; tat protein; effector protein; transduction domain; HopPcoA protein.  
XX Pseudomonas syringae.  
OS  
XX WO200119393-A1.  
PN  
XX PD  
XX 22-MAR-2001.  
XX  
XX 13-SEP-2000; 2000WO-US24977.  
XX  
XX 13-SEP-1999; 99US-0153507.  
XX  
XX (CDRR ) CORNELL RES FOUND INC.  
XX Colimer A, Beer SV;  
PI  
XX WPI: 2001-257850/26.  
XX N-PSDB; AAF55683.  
DR  
XX Delivering effector proteins into target cell for use in protein  
PT therapy, involves introducing effector protein fused to protein  
PT transduction domain of human immunodeficiency virus Tat protein, into  
PT target cells -  
XX  
XX Example 2; Page 23; 43pp; English.  
XX  
XX The present sequence represents a HopPcoA protein of Pseudomonas syringae  
CC pv. tomato DC300 CEL.. This is an effector protein, which is used in  
CC the method of the invention. The specification describes a method for  
CC delivering effector proteins into a target cell. The method comprises  
CC introducing an effector protein fused to a protein transduction domain  
CC of a human immunodeficiency virus (HIV) tat protein into the target  
CC cell. The method is used for delivering a heterologous effector protein  
CC such as an effector protein produced by bacterial plant pathogen, animal  
CC pathogen or a rhizosphere bacteria, or a protein secreted and/or  
CC delivered into eukaryotic cells by a Type III secretion system or a  
CC hypersensitive response elicitor, an avirulence protein, a  
CC hypersensitive response and pathogenicity-dependent outer protein, a  
CC virulence protein or a pathogenicity protein, into an eukaryotic cell.  
CC The method is useful for delivering effector proteins for use in  
CC pharmaceutical, insecticidal, fungicide, herbicide and other  
CC applications.  
XX  
XX Sequence 486 AA;  
SQ  
Query Match 77.7%; Score 1869.5; DB 22; Length 486;  
Best Local Similarity 78.4%; Pred. No. 3.8e-147;  
Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;  
QY 1 MHINQSAQPPGVA MESFRTASDAS LASSSVRSVSTTSCRDQAITDY LKHHVFAAHRFS 60  
| | | | | : | | | | | | | | | | | : | | | | | | | | | |  
DB 1 MHINRRVQGPPTATDSFRTASDAS LASSSVRSVSSDQREINAIADY LTHVFAAHLKP 60  
QY 61 VIGSPDERDALANECIDALVETRAARKLYSEGEPATIAETAKAEFDLATITSSNAF 120  
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 61 PADSADGQAAVVDHNAQITALETRASRIHFEGEPATIAETPAKAEKDLATITTSNAL 120

QY 121 ENTPEAASVLQTMQPAINKGDMILATPLKPLPLISGALS GAMDQVGTMMQBARAGDLHY 180  
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 121 RATPEFAMASLLQYMQPAINKGDMILAPLKPPLISGALSGAMDQVGTMMQBARAGDLHY 180  
QY 181 LSTSPDKLHDMASVYKRHSFALGRQVDMGI AVQTESALNVVRTVLAPALASRPSVGA 240  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 181 LSNASPDRLHDMASVYKRHSFSLARQVLTGTGVAQVTSARNAVKTVLAAPALASRAVGA 240  
QY 241 VDFGVSTAGLVNAGSGPRMLSVQSRDQLRGAFVYLGKKDKPEKALSEETDMIDAVKA 300  
| | | | | : | | | | | : | | | | | | | | | | | | | | | | | | | | | |  
DB 241 VDLGVSYAGGLAANAGFGRRLISVQSRHQRCQALVYGLKDKPEKQQLSEENDMLEAVKA 300  
QY 301 IKSASVSGAALNAGKRMAGLPDYATDGLKAVRSVSAISLTIRGALAGGVAGVSKQK 360  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 301 IKSASVSGAALNAGKRMAGLPDYATDGMAGVRSVSAISLTIRGALAGGVAGVSKQK 360  
QY 361 MATRNTDTSATKAASQLSNVGSGVGFQMTTGLATDPAVKKAEFTIDVKSTASST 420  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 361 MATRNTDTPATKAASQLTNLGSAALVAFQMTTALTTDPAVKKAEFTIDVKSTASST 420  
QY 421 TSYVADQTVKLAKTVKIDSGEALISYTGASLRSTVNNLRHRSAPREADI EEGGISAERSSET 480  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 421 TGVVADQTVKLAKTVKIDSGEALITHTGASLRSTVNNLRQRPAREADI EEGGTAA-SPEEI 479  
QY 481 PQLRR 486  
| : |  
DB 480 PFRPKR 485  
RESULT 4  
AAE20110  
ID AAE20110 standard; Protein; 1463 AA.  
XX  
XX AAE20110;  
AC  
XX DT 18-JUN-2002 (first entry)  
XX  
XX DE Lactobacillus rhamnosus outer membrane protein rompA.  
XX  
XX Enzyme; flavour; aroma; texture; nutritional; dairy manufacture; therapy;  
KW fermentation process; anti-infection; rotavirus infection; heart disease;  
KW infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder;  
KW anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori;  
KW antihypertensive effect; urogenital infection; hepatic encephalopathy;  
KW bowel syndrome; endocarditis; transgenic microbes; outer membrane protein;  
rompA.  
XX  
XX Lactobacillus rhamnosus HN001.  
OS  
XX WO200212506-A1.  
PN  
XX 14-FEB-2002.  
PD  
XX 08-AUG-2001; 2001WO-NZ00160.  
PF  
XX 08-AUG-2000; 2000US-0634238.  
PR 28-NOV-2000; 2000US-0724623.

XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA (VIAL-) VIALACTIA BIOSCIENCE NZ LTD.  
XX  
PI Glenn M, Havukkala J, Blokberg JN, Lubbers MW, Dekker J,  
PI Christensen AC, Holland R, O'cole PW, Reid JR, Coolbear T;  
XX  
DR WPI: 2002-241760/29.  
XX N-PSDB: AAD31881.  
PT New polynucleotides and polypeptides from Lactobacillus rhamnosus,  
PT useful in e.g. improving the flavor, aroma, texture and health-related  
PT benefits of milk-derived products, or in increasing properties of  
PT microbes -  
XX  
PS Claim 2; Fig 64; 257pp; English.  
XX  
CC The present invention relates to a new isolated polynucleotide comprising  
CC a sequence present in Lactobacillus rhamnosus strain HN001 and encoding a  
CC polypeptide capable of modifying the flavour, aroma, texture, nutritional  
CC and health benefits of milk-derived products, and/or survivability of  
CC microbes in dairy manufacturing processes. The polynucleotides are useful  
CC for improving the properties of microbes used in the manufacture of milk-  
CC derived products such as cheeses, yogurt, fermented milk products, sour  
CC milks and buttermilk; in modifying the flavour, aroma, texture and health  
CC -related benefits of milk-derived products and in increasing the survival  
CC of microbes during industrial fermentation processes. The bacteria may be  
CC used to increase resistance to enteric pathogens and anti-infection  
CC activity, including treatment of rotavirus infection and anti-infection  
CC diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis;  
CC liver cancer reduction; reduction of small bowel bacterial overgrowth;  
CC immune system modulation and treatment of autoimmune disorders and  
CC allergies; treatment of allergic responses to foods; reduction of blood  
CC lipids and prevention of heart disease; antihypertensive effect;  
CC prevention and treatment of urogenital infections, Helicobacter pylori,  
CC or hepatic encephalopathy; treatment of inflammatory bowel disorder and  
CC irritable bowel syndrome; modulation of endocarditis; and for improved  
CC protein and carbohydrate utilization and conversion. The transgenic  
CC microbial population can be administered to a mammal as an anti-  
CC carcinogenic agent. The present sequence is Lactobacillus rhamnosus  
CC outer membrane protein rompa.  
XX  
SQ Sequence 1463 AA;  
Query Match 6.1%; Score 147; DB 23; Length 1463;  
Best Local Similarity 20.8%; Pred. No. 0.0058;  
Matches 104; Conservative 64; Mismatches 241; Indels 92; Gaps 12;  
QY 1 MHINSAQQPGVWMSFRTASDASLSSVSSTSCRDQATITVLRHVFPAARFS 60  
Db 714 INNNSADVTASQAKATSAASDAVSASEOSTAGSHADNNE----ITSLASDAEKQS 768  
QY 61 VIGSPDERDALAHNEQIDALVETFRANRLYSEGETPATI-----AETFAKAEFRLATT 115  
Db 769 QIALAASKAASAASSAAIAIV--ASSAASEASAAAVSNADSNASAAAYASVASE 826  
QY 116 ASSA-----FENTPFAAASVLTQMQPAINKGDMATPLKPLTPLISGALSGAMDVG 167

Db 827 ASASAANSSGGATASFASAAAM-----SALSTAQVAAK 865  
QY 168 TRQMDRARGDLHYLSTSPDKLHDANA/VKRRHPALGRQVDMGIAVQFFSALNVVR--T 225  
Db 866 VAVSDMAAAG-----SAAVAASAAQSDSNKQATRAATASQAALDDLYKITSLT 913  
QY 226 VLAAPALASRPVQG-----AVDFGVSTAGGLVANNAGFGRMLSVQSRDQLNGGA 274  
Db 914 DVASGASSASAEAGQASTATSAVASASSSASEGSGVAHQAGSSASPDVQSGSAQHAS 973  
QY 275 FVLGKQKEPKAALSEETDMLDYKTKISASYSQAALNAGRMNGLELDVATDGLKAVRS 334  
Db 974 TAAASASSYPRKDSGIQSLASQAASEAFAASSASAASTAS---AAVGFSAASDASEQART 1029  
QY 335 LVASATSLRNGLAGVAGVSKIQMAATKNTDTSATKAASQLSNLVGSGVFAGWTTA 394  
Db 1030 AASADVYASSAASSTNSVSAAM-----SATKAGDSKAA-----AGFSSA 1069  
QY 395 GLATPDPAVKRAESFIQDKVSTAS--STTSYVADQTVKLATVKIKMSGEAISSTGASLNS 452  
Db 1070 ASNAASSAKAEAVASEAASADSDVYASSAASAAAGFDRKASAEAGAAASAAASAS 1129  
QY 453 TVNNLRHRSAPAEADIEEGGIS 473  
Db 1130 AAAQTRGAGSSASEAQAS 1150  
RESULT 5  
ABU08784  
ID ABU08784 standard; Protein: 2122 AA.  
AC ABU08784;  
XX  
DT 28-MAY-2003 (first entry)  
XX  
DE Moraxella catarrhalis outer membrane protein, OMP106.  
XX  
KW Outer membrane protein-106; OMP106; Moraxella catarrhalis infection;  
KW vaccine.  
XX  
OS Moraxella catarrhalis.  
XX  
FH Key  
FT Peptide 1..68 Location/Qualifiers  
FT /label= Signal\_sequence  
FT Protein 69..2122  
FT /label= Mature\_OMP106  
FT Region 69..111  
FT /label= N-terminal sequence of mature OMP106  
FT /note= "Specifically claimed in claim 9"  
FT 779..880  
FT /note= "Encoded by ATCTAAGCC"  
XX  
PN US2002177200-A1.  
XX  
PD 28-NOV-2002.





CC repeat domain; and (d) a carboxy terminal domain. The extracellular amino  
 CC terminal domain comprises 5 genomic exons, the amino terminal extension  
 CC comprises 4 genomic exons, each repeat unit comprises 5 genomic  
 CC exons and the carboxy terminal domain comprises a transmembrane anchor  
 CC with a short cytoplasmic domain, and further comprises 9 genomic  
 CC exons. The gene for CA125 is located on human chromosome 19q 13.2.  
 CC Also included are isolated CA125 repeat domains, nucleic acids  
 CC (including variants, homologues and degenerate versions) encoding CA125  
 CC proteins or repeat units, a vector comprising the nucleic acid, a  
 CC cultured cell comprising the vector, a method of expressing CA125 antigen  
 CC in a cell, the amino acid sequences of the CA125 repeat units (or  
 CC their variants, fragments or sequences 50% identical to them),  
 CC a purified antibody that selectively binds to an epitope for detecting and  
 CC receptor-binding domain of CA125 protein, a diagnostic for detecting and  
 CC monitoring the presence of CA125 antigen (comprising recombinant CA125  
 CC having at least one repeat unit of the CA125 repeat domain including  
 CC epitope binding sites), a therapeutic vaccine to treat mammals with  
 CC elevated CA125 antigen levels or at risk of developing a disease or  
 CC disease recurrence associated with elevated CA125 antigen levels  
 CC (comprising recombinant CA125 repeat domains including epitope binding  
 CC sites) and an antisense oligonucleotide that inhibits the expression of  
 CC CA125. The CA125 molecule, particularly the multiple repeat domains are  
 CC useful as a gold standard for detecting and monitoring the presence of  
 CC CA125 antigen, which can be used for diagnosing, monitoring or treating  
 CC patients with ovarian cancer and other carcinomas where CA125 is  
 CC expressed. The molecules are also useful for developing a vaccine  
 CC against cancer. The present sequence is a CA125 repeat protein.  
 CC  
 XX  
 SQ Sequence 10431 AA;

Query Match 5.9%; Score 143; DB 24; Length 10431;  
 Best Local Similarity 19.0%; Pred. No. 0.2;  
 Matches 113; Conservative 87; Mismatches 190; Indels 204; Gaps 25;  
 QY 3 INGSAGPPGVAWESFRTPASDASLASSVSYS-----TSCSDLAQI-----TDYLKHHVF 54  
 Db 4306 VNOGLQDE-----NSTSQVASSKTTSTVTTHVSGDATHVKTQATFSSGTSISSPHQF 4360  
 QY 55 --AAHNFVIGSPDERDALAHNEQIDALVETRANRLYSFG-----ETPATI 99  
 Db 4361 ITSTNFTFDVSTNPSTSLIMTESSGVTITTOGTGAAGOGYLLDSTWPLYETFLPAAV 4420  
 QY 100 AETFAKAKEFDRLA-----TTASAFEN-----TPAAASVLYQMPALIN--KCDW 143  
 Db 4421 TPDFMQSEKRTLLISKGPKDVWTSPSPVAETSPSSLTPELVYTI----PPAISFLAQGH 4476  
 QY 144 LATPLKPLTPLISGALSGAMDOVGTQAMPDRAGDHLVLTSPDKLHD-----AAAV 194  
 Db 4477 TSSPYVATSVLTSGLV-----KTDMLNTSMERVTNSQNINPSELIATLATIT 4527  
 QY 195 SVKRHSPLAGQVVDGIAVOTFSALNVVRYLAPALASRPSVQAVDFGVSTAGGLVAN 254  
 Db 4528 DIETIHPSINKAVYNNG---TASSAHVLSHSL--PVSSBPS-----TATSPMVA 4572  
 QY 255 AGFGDMLSVQSRDLRGAVFLGAKDKPEKAAALSETDMLANVYALKSAYSGAALNAG 314  
 Db 4573 SSMGALASIS-----IPG-----SETTD----- 4591

QY 315 KRWAGLPIDVATDGLKAVRSIVSATSLIKGLALAGYAG--VSKLQKM----- 361  
 Db 4592 --IEGEPTSLTAGRKENSITLQENNSTEENIILSVSVGAITEATKREVPSPDAFTPT 4649  
 QY 362 ---ARK-----NITDSATKXAYSQL-----SNLVGSYG-- 386  
 Db 4650 PAQSTKFPDIFSVASSRLSNSPPMTISTHMTTQTQSSGAVSKIPALDSTLETSAGRP 4709  
 QY 387 --VFAGW-----TTAGLATDPAVKRAESFIQDKVKSTAS-----S 419  
 Db 4710 SVTEGFRHAKKITTMANNIDVXQVSOINPPODASSPSQAPVLYVTTLPSSVATPQGMHS 4769  
 QY 420 TTSYVADQYKIAKTVK----DMSGEAISSGASLIRSTVNNLKRHSAPADIE 468  
 Db 4770 TSSPVMSVLTSSIVTKAGKVDTSLETVTSSPQSWNTLDDISVTSATTDIE 4823  
 RESULT 8  
 AAW98149  
 ID AAW98149 standard; Protein; 2285 AA.  
 XX  
 AC AAW98149;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE Bacillus subtilis metalloprotease YOMI.  
 XX  
 KW YOMI; metalloprotease; protease; textile; animal feed; detergent;  
 KW Gram-positive bacterium.  
 XX  
 OS Bacillus subtilis.  
 XX  
 FN W03914342-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 08-SEP-1998; 98WO-US18828.  
 XX  
 PR 15-SEP-1997; 97GB-0019636.  
 XX  
 PA (GENM/ ) GENENCDR INT BV.  
 PA (GENM/ ) GENENCDR INT INC.  
 PI  
 PI Estell DA;  
 XX  
 DR WPI; 1999-229541/19.  
 DR N-PSDB; AAX24980.  
 XX  
 PT Compositions containing metalloprotease from Gram-positive  
 PT microorganism  
 PS  
 PS Claim 4; Fig 1A-O; 59pp; English.  
 XX  
 CC The present sequence is metalloprotease YOMI of Bacillus subtilis.  
 CC The invention relates to the discovery of this previously unknown  
 CC metalloprotease, uses of the metalloprotease in industrial and



CC agricultural applications, and advantageous strain improvements  
CC based on genetically engineering a Gram-positive microorganism to  
CC delete, underexpress or overexpress the enzyme. Due to overall  
CC relatedness of the enzyme with Pseudomonas lasa protein, YOMI  
CC appears to be a member of the M23 metalloprotease family. The  
CC metalloprotease can be used in claimed cleaning compositions. It may  
CC also be used for peptide hydrolysis, waste treatment and for  
CC animal feed and compositions for the treatment of textiles. It may  
CC also be used for peptide hydrolysis, waste treatment and for  
CC cleaving recombinant fusion proteins. Expression vectors  
CC comprising a nucleic acid (see AXX24980) encoding the metalloprotease  
CC and host cells are claimed. Gram-positive cells in which YOMI is  
CC inactivated (by gene mutation or deletion) are used for production  
CC of heterologous proteins, especially enzymes, hormones, growth  
CC factors and cytokines.

XX  
XX  
SQ Sequence 2285 AA;

Query Match 5.8%; Score 140; DB 20; Length 2285;  
Best Local Similarity 20.5%; Pred. No. 0.041;  
Matches 128; Conservative 80; Mismatches 224; Indels 192; Gaps 26;

QY 15 MESFRTASDASLASSTSVSTSTSCRDQAITDYLRKHVFAAHRSY-IGSPDERDAALA 73  
DB 251 IEIQQOQOQVONQNTNRGSSMOSNRQAVDYTLN----ANSLNVTGSNNITSQIQS 306  
QY 74 HNEQIDAL---VETRANRLYSEGETPATIAETP-----AKA 106  
DB 307 LNNQFRELASNAQTANQSSFG--AEITQTFKSMSTYLLISGLFYAIGSLKPEMSQA 363  
QY 107 EKFDRLATTSASAFENTPPAAASYLYQNPALNKGDMPLPLKPLTLISGALSANDQY 166  
DB 364 IEIDTLMTNIRRVNNEPDYKNELL---QESIDLDGDTLSNKITDILQMTGDFGRMGFDES 420  
QY 167 GTQMDDRARGDILYLS-TSPDKLHDMAVSVKRSPALGRV-----VDMGIAVQTF 217  
DB 421 ELSTLTKTQVLYONVSDLTTPDYNTLTAAKLNTNIAANDSISADKNLVDMNVAATTL 480  
QY 218 SALNVRTVLAIPALASRPSVQAVDFGVSTA-----GGLVAN-----AGFGDRMLS 263  
DB 481 DLANSIRK--AGSTASTFGVEINDLIGTTALASTTRESGNIVGSLKTFARIGNQSS 538  
QY 264 VQSRQQLGAGFVLRKKQKEPKALSP---ETDNL-DAYKAIKKSASYSG----- 308  
DB 539 IKALRQIGISVKTAGGEAKSASDLISEVAGKWDTLSDAQKNTSIGVAGIYQLSRVAMM 598  
QY 309 ----AALNAGK----- 315  
DB 599 NNFSTIAQNAAKTAANSTGSANSEQKKAADSLQARVKKLQNNTEFAITASDAFISDGLIE 658  
QY 316 --RMAGLPUDVATDGLKAV--RSLVSATIS-----LTKGLALAGG-VAGVSKL--QDMA 362  
DB 659 FTQAAGSLNASTGVIKSVGFPLPLAAVSTATLLISKNTRTLASLILGTRAMQOERTLA 718  
QY 363 TKNITDSATKAASQ-----LSNLVGSVGVFAQMTTAGLATDPA-VKKA----- 405  
DB 719 TAGLEAGMTAAVAASRVLTALRGLLVSTLVGGAFAALGMALESIISSFAEAKKAKDPE 778

QY 406 -----ESFIQ-----DRKYSTASTSYVAQQ---IVKIAKT-----VK 436  
DB 779 QSQOQTNVEAITTNKDSYDNLQYKELQVRESRSUTSDEBOXYLOVYQQLAOTFPALVK 838  
QY 437 --DMSGAISSTGASLSTVNNLR 458  
DB 839 GYDSQGNALIKTNKELEKALENTK 862

RESULT 9

ABB71736  
ID ABB71736 standard; Protein; 1229 AA.  
AC ABB71736;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 42000.  
XX  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
EN W0200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li FWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
XX N-PSDB; ABL15839.  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
PS Disclosure; SEQ ID NO 42000; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences and the encoded proteins  
CC sequences (ABL1840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX	Sequence	1229 AA;	
SQ	5.7%; Score 138; DB 22; Length 1229;		
	Query Match	Best Local Similarity 20.9%; Pred. No. 0.025;	
	Matches 107; Conservative 73; Mismatches 205; Indels 128; Gaps 21;		
QY	11 PGVAMESFRITASDASLASS--VRSVSTISGRDQALITDY LKHVFAAHFSEVIGSPDER 68		
DB	166 PGITVATYTS PGVSYSSASPGISKVATYSSPSLSVYPYAPVTSVEVYS--SP--- 219		
QY	69 DAALAHNEQID--ALVETRAARLYSSEGETPATI-----AETPAAEKFD--- 110		
DB	220 --AYTYSKTTPGYSKVEYSSPGYSYSGQISPGISRATYSPSVSYSAPTIAKVTSYAPS 277		
QY	111 -RLATTAS-SAFENTPEFAA---SVLQYKQPAINKGDMLATPLKPLTPLIS-----GA 158		
DB	278 VKIATTSLSLSHGTCYSASVAPSTIKYQSVDVSHQYISKRYAAKPAITVNAASYGGT 337		
QY	159 LSGANDQ-----VGTKMDBARGDILHLYLSTP--DKLHDAMAVSVKRHSPPALGRQVDM 210		
DB	338 ASGALS HQYVSGPAIAKVSTYAAPVATYSSGPALSKLSTSYGAS--GSGAVSHQYYSK 394		
QY	211 GIAVQTFALNVVRYTLAPALAS---RPSVQGAIDRCVSTAGGLVANAAGCGRHLSTVQSR 267		
DB	395 PAVAIAPAVAKATYAAPAISSYSTGPALSKVASVAAPTVSTVSSGYGVS----- 446		
QY	268 DOLRGAEVLGKKDEPKALSEETDMLDAVKAI-KSASVSGAALNKGKRMAGLPL----- 322		
DB	447 ----SSGAVSHQYVSKPAVALS-----AAPAIKAKATYAAPALIST--YAAAPVTVKX 493		
QY	323 -----DVATDGLKAVRSVLVSATSLTNKGLAAGYAGV 355		
DB	494 ATGYGSGSGYSSGAVSHQYVSKPAVAKVATYAAPALSTYSAAAPATVKIATSYGSGHGA 553		
QY	356 SKIQKATNITDSATKAAVQSLNLVGSVGFENGRTYGLATDPANAKAESFIODKVKX 415		
DB	554 VSHQYVSKPAVALSAAFA-----IAKVATYASPAISTYATAPVVSKVATYAAPSIAT 605		
QY	416 TASSTTSYVADQTVKLAKT---VKDMSGELIS 444		
DB	606 YSSAPF-----LAKVYSQAADVSHQYIS 629		
RESULT 10			
AAB76536			
ID	AAB76536 standard; Protein; 1277 AA.		
XX	AAB76536;		
XX	11-APR-2001 (first entry)		
XX	Corynebacterium glutamicum MCT protein SEQ ID NO:54.		
DE	Corynebacterium glutamicum MCT protein SEQ ID NO:54.		
KW	Corynebacterium glutamicum; brevibacterium lactofermentum; MCT; membrane construction and membrane transport protein; petroleum spill;		
KW	hydrocarbon degradation; gram positive aerobic bacterium; marker;		
PT	Identification; microorganism; fine chemical production; transformation;		
XX	genome mapping; genetic engineering.		
OS	Corynebacterium glutamicum.		
XX	W0200100805-A2.		
XX	04-JAN-2001.		
PD	23-JUN-2000; 2000WO-IB00926.		
PF	25-JUN-1999; 99US-0141031.		
XX	08-JUL-1999; 99DE-1031454.		
PR	08-JUL-1999; 99DE-1031478.		
PR	08-JUL-1999; 99DE-1031563.		
PR	09-JUL-1999; 99DE-1032122.		
PR	09-JUL-1999; 99DE-1032124.		
PR	09-JUL-1999; 99DE-1032125.		
PR	09-JUL-1999; 99DE-1032128.		
PR	09-JUL-1999; 99DE-1032180.		
PR	09-JUL-1999; 99DE-1032182.		
PR	09-JUL-1999; 99DE-1032190.		
PR	09-JUL-1999; 99DE-1032191.		
PR	09-JUL-1999; 99DE-1032209.		
PR	09-JUL-1999; 99DE-1032212.		
PR	09-JUL-1999; 99DE-1032227.		
PR	09-JUL-1999; 99DE-1032228.		
PR	09-JUL-1999; 99DE-1032229.		
PR	09-JUL-1999; 99DE-1032230.		
PR	14-JUL-1999; 99DE-1032927.		
PR	14-JUL-1999; 99DE-1033005.		
PR	14-JUL-1999; 99DE-1033006.		
PR	27-AUG-1999; 99DE-1040764.		
PR	27-AUG-1999; 99DE-1040765.		
PR	27-AUG-1999; 99DE-1040766.		
PR	27-AUG-1999; 99DE-1040830.		
PR	27-AUG-1999; 99DE-1040831.		
PR	27-AUG-1999; 99DE-1040832.		
PR	27-AUG-1999; 99DE-1040833.		
PR	31-AUG-1999; 99DE-1041378.		
PR	31-AUG-1999; 99DE-1041379.		
PR	31-AUG-1999; 99DE-1041395.		
PR	03-SEP-1999; 99DE-1042077.		
PR	03-SEP-1999; 99DE-1042078.		
PR	03-SEP-1999; 99DE-1042079.		
PR	03-SEP-1999; 99DE-1042088.		
XX	(BADI ) BASF AG.		
PA	Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;		
PI	WPI; 2001-071486/08.		
XX	N-PSDB; AAF67769.		
DR	Corynebacterium glutamicum nucleic acids encoding membrane construction		
PT	and membrane transport proteins or their portions, useful for typing or		
PT	identifying C. glutamicum or related bacteria, and as markers for		

PT transformation -  
XX  
PS Claim 20; Page 242-246; 1119pp; English.  
XX

CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane  
CC construction and membrane transport (MCT) proteins given in AAB76510 to  
CC AAB76547. The MCT nucleic acids and proteins are useful in the  
CC identification of microorganisms which can be used to produce fine  
CC chemicals, for modulating fine chemical production in C. glutamicum or  
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
CC identification of C. glutamicum or related bacteria, as reference points  
CC for mapping C. glutamicum genome, and as markers for transformation.  
CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
CC example from the present invention.  
XX

SQ Sequence 1277 AA;  
Query Match 5.7%; Score 138; DB 22; Length 1277;  
Best Local Similarity 21.7%; Pred. No. 0.027;  
Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;

QY 7 AQQPPGVAMESFRTASDASLSSVSRSVSTSCRDQATDYIKHHVFAAHFVSIGSPD 66  
Db 313 AHTAPDVLGVAMPFAVFAVSAVPTGDSASVVEGMSLVHLEHHI-----VLASDV 365  
QY 67 ERDAALAHNEQIDALVETRANRLY-----SEGETPATIAETFA-KAERTDRLATYAS 117  
Db 366 PTGALIKVSATADEVVDRLGLVIVRAIEADAEQNLIAIEAEFALRGKNAVAARNT 425  
QY 118 SAFE---NTPPAAASVLOQMOP-----AIKKGMLATPLRLPL-----ISGALSGMD 164  
Db 426 SALPTTVTPPSARAATVVAPESMRPFVVISGD-----RNPINVSDVAASTLAGLPG 477  
QY 165 QVGTQMDRARGDLHYISTSPDKLIDAAVSVKHH-----SPALGRQVYDMGI---AVQT 216  
Db 478 VIVHGAMTSAIGELIAGAFDEQIQTPAKKVEYATYTLAPLPEGEIEFVSERAVDN 537  
QY 217 FSAALVVRVTLAPALASRPVQGVAGVSTAGLVA--NAGFDRMLSVQSRDQLRGA 274  
Db 538 RPKMGEVRYTYA-----TYNGNL---VLTATVVAADSTFYAFPOGQIOSG----- 581  
QY 275 FVLGKKDEPKAAISFETDMIDAKIKASASYSGAALNAGKRMAGLPIDVATDGLKAVR- 333  
Db 582 --MGHEARRNSQAAAIWDRAADATRNK-----LGFSTVEIVNNPREVTVAAGEKFFHP 633  
QY 334 -SLVSATSLTNGIALAGVAGVSKIQKKAATNTTSDATKAASQLSNLVGSVGFAGMT 392  
Db 634 DGVLYLTQFTQVGHATL-GVAQIAEKREAHALNQRAIFAGHSVGEYNAALAAAGVLSLES 692  
QY 393 T-----AGLATDPAVKAESFIQ-----DVKSTASSTTSVVAQDTVKLAKTVKD 437  
Db 693 VLEIYVRGLTMRVLVDRENGLSYALAAALRNKGLTADIVFDVA-----SVSE 744  
QY 438 MSGE-----AISTGASLSTVNNLRHRSAPEDIEF--GGISAF-----SRS 478  
Db 745 ASGEFLIIVNNLAGLQYAVAGTQAGL-----AALRADVENRPAQCAFILIPGI 794

QY 479 ETPQLRL 487  
Db 795 DVPFHSSKL 803

RESULT 11  
AAB76534  
ID AAB76534 standard; Protein; 2012 AA.  
XX  
AC AAB76534;  
XX  
DT 11-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MCT protein SEQ ID NO:50.  
XX  
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;  
KW membrane construction and membrane transport protein; Petroleum spill;  
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;  
KW identification; microorganism; fine chemical production; transformation;  
KW genome mapping; genetic engineering.  
XX

OS Corynebacterium glutamicum.  
XX  
PN W0200100805-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-1B00926.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 08-JUL-1999; 99DE-1031454.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031563.  
PR 09-JUL-1999; 99DE-1032122.  
PR 09-JUL-1999; 99DE-1032124.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032180.  
PR 09-JUL-1999; 99DE-1032182.  
PR 09-JUL-1999; 99DE-1032190.  
PR 09-JUL-1999; 99DE-1032191.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032212.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 09-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032927.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040830.  
PR 27-AUG-1999; 99DE-1040831.  
PR 27-AUG-1999; 99DE-1040832.  
PR 27-AUG-1999; 99DE-1040833.

```

Db      1220  RPGRGGEIRRVYA-----TNGNL---VLATLVVAASFTFAFPQGGIQSGQ----- 1265
QY      275  FVLGKDKDEPKALSEETIMDAIKAKISASTSGALLNGKRAALGPDIYATDGLKAVR- 333
      1111  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1264  --MKMERARNISQAARAIMWRADAAHTNRK-----LGFSTVELIYNNRRETYAAGEKEFHH 1315
QY      334  --SLVSAITSILTRKIGLALAGGAGVSKLQIRAKATKNITDCAKVAQSLSNLIVGVSFGAGMT 392
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1316  DQGLYLTLQTYQGNATL-GVQIAIEBEBHALLNQRATFAGHSVGBENYALLAAVAGVLTLES 1378
QY      393  T-----AGATDPAVKAESFIQ-----DKVSTASTTSIVYADQTVKALAKTVKD 437
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1375  VLEIIVYKRGELTWRLVDRDENGLSNYALLALPFRKQGLFADINVDYVA-----SVSE 1420
QY      438  MSQG-----AISTGASIRSTVNNLRHRSAPAEADIEE--GGISAF-----SRG 478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1427  ASGGELEIYNNVNLGAGQYVAVNOTQAGL-----ALLRADVENNAPQGNALFILRGI 1478
QY      479  ETPQLRL 487
      : : : : :
Db      1477  DVPEHSSKL 1485

```

XX	RESULT 12
XX	AA692485
ID	AA692485 standard; Protein; 2993 AA.
XX	
XX	AA692485;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	C glutamicum protein fragment SEQ ID NO: 6239.
XX	
KW	Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW	organic acid synthesis.
XX	
OS	Corynebacterium glutamicum.
XX	
PN	EP1108790-A2.
XX	
PD	20-JUN-2001.
XX	
PF	18-DEC-2000; 2000EP-0127688.
XX	
PR	16-DEC-1999; 99JP-0377484.
PR	07-APR-2000; 2000JP-0159162.
PR	03-AUG-2000; 2000JP-0280988.
XX	
XX	( KYOWA ) KYOWA HAKKO KOGYO KK.
XX	
XX	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	
DR	WPI; 2001-376931/40.
DR	N-PSDB; AAH67704.
XX	
DT	Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT -  
XX  
PS Claim 17, SEQ ID NO: 6239; 246pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from corynebacterium bacterium, and identifying a homologue of a gene derived  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 2993 AA;  
Query Match 5.7%; Score 138; DB 22; Length 2993;  
Best Local Similarity 21.7%; Pred. No. 0.089; Indels 126; Gaps 23;  
Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;  
QY 7 AQQPGVMEFRTASDASLASSSVRSVSTTSCDLOATDYLKHFVAFHRSVYSGPD 66  
DB 1025 AHTAPDVLGRAPVAFVAVKSAVITPGTDSASVVEGMLSVLHSHH-----VLKSDV 1077  
QY 67 ERDALLAHNEQIDALVETRANRLY-----SEGETPATIAETFA-KAEKEDRLATTAS 117  
DB 1078 PTDGALKVSAIADAEVDTDLGRVIVRAEIDAEGLNLIATIAERFALRGKGNVARTNT 1137  
QY 118 SAFE---NTPEAAASVLYQMP-----AIKKGPMIATPLKRLPL-----ISGALSGMD 164  
DB 1138 SALPTTVDTPRSRAVATVAPESMRFVVISGD-----RNPINSVAAASLALGPG 1189  
QY 165 QVGTQMDRARGDLHYLISTSPDKLIDMAVSVKRN-----SPALGRQVDMGI---AVQT 216  
DB 1190 VIVHGAWTSAIGELLAGAFAFDEQITPAKAVVEYATYTLAPVLPGEIHFSEVRSADVNI 1249  
QY 217 FSNLVAVTTLAPALASRPSVQGVDFGVSTAGLVA--NAGFDRLISVQSRDYLKGA 274  
DB 1250 RPKAGEVTVTA-----TVNGNL---VLATATVAAAPSTYAFPGQIGSQG----- 1293  
QY 275 FVLGKKDEPKALSEETDMLDAYKAIKGSASVSGALNGRMAGLPUDVATDGLKAVR- 333  
DB 1294 --MGHEARNSQAARAIMDRADAHTRNK-----LGFSTVEIVENPREVTVAGKFFHP 1345  
QY 334 -SLVASTSLTKNGIALAGVAGVSKLQKAKTNTTDSATKAAVSQSLNVGSGVFAGMT 392  
DB 1366 DGVLYLTQFTQVGWATL-GVAQIAEREAHLNQRAPVFGHSGEVNLAALAAVGLSLES 1404  
QY 393 T-----AGLATDPAVKKAESFTQ-----DKVSTASTTSYVADQTVKALKYKD 437  
DB 1405 VLEIYVRGLTMRILVDRDENGSLNYALALRPKMGTLADNVEYVA-----SVSE 1456

QY 438 MSGE-----AISTGASLRSTVNNLHRSPADIEE--GGISAF---SRS 478  
DB 1457 ASGEFLEIVNNYLAGVAVAGTQAGL-----AALRAVDENRAPCGRAFTILPGI 1506  
QY 479 ERPFQRLRL 487  
DB 1507 DVPFHSSKL 1515  
RESULT 13  
AAR48993  
ID AAR48993 standard; Protein: 1026 AA.  
XX  
AC AAR48993;  
XX  
DT 14-SEP-1994 (first entry)  
XX  
DE rsaA S-lyase protein.  
XX  
KW C. crescentus; rsaA; paracrystalline; S-layer; protein; heterologous;  
KW cellulase; xylase; metallochionein; restriction site;  
KW reading frame; fusion protein; bioreactor; toxic metal; sewage;  
KW waste water; wood pulp suspension; cell surface; vaccine; fish.  
OS Caulobacter crescentus.  
XX  
PN CA2090549-A.  
XX  
PD 10-DEC-1993.  
XX  
PF 26-FEB-1993; 93CA-2090549.  
XX  
PR 09-JUN-1992; 92US-0895367.  
XX  
PA (UYBR-) UNIV BRITISH COLUMBIA.  
XX  
PI Bingle WH, Smit J;  
XX  
DR WPI; 1994-066249/09.  
DR N-PSDB; AA057972.  
XX  
PT Prod. of heterologous polypeptides in bacteria, partic.  
PT Caulobacter - by expression of a fusion prod. of the polypeptide  
PT sequence and a bacterial S-layer protein gene  
XX  
PS Claim 17; Fig 6; 27pp; English.  
XX  
CC This sequence is encoded by the C. crescentus rsaA gene and represents  
CC the paracrystalline S-layer protein. The rsaA gene was used in the  
CC production of the heterologous protein of the invention. The  
CC heterologous protein is produced by cloning a polypeptide coding  
CC sequence, eg. cellulase, xylase or a metallochionein, into a  
CC restriction site within the rsaA gene which preserves the rsaA reading  
CC frame and expressing the fusion sequence in Caulobacter. This S-  
CC layer protein bacterial system can be used in bioreactors, eg. to  
CC bind toxic metals in sewage waste water etc. or for the treatment of

**SQ Sequence 1026 AA;**

Query Match	5.78;	Score 136;	DB 15;	Length 1026;

Matches 102; Conservative 75; Mismatches 165; Indels 140; Gaps 23;

```

QY      22 SDASLASSVRSVSTTSCRDQAITDYLKHHFAAHFVSIGSPDEDAALAHNEQIDAL 81
      ||| : :::: :| : | : |
Db      42 SDAALTNVLKLNSTFAVAIQ-----YQFTGVAPSAG-----LDFL 81

```

```
QY      82 VETRAN-----RLYSGETPATIAEIPAKAEKF-----DLATTTASSAENTPPAAA---- 128
        ||: | ||: ||: ||: | ||:
Db      82 VSDTNNNDINDAIYSK-----FAQENRFINSINLATGAGAG--ATAFAAAATG 129
```

```

Qy      129 -SYLQYQPAINK--GWLATPLKPLPLISGALSAMDQYGTMMEDRAGEDLHYISTSP 18
      | | : | : | : || : : : || :
Db      130 VSAQYVATAYDKITIGNAVAT-----AACDVAAVAFLSRQNIIDYLT--- 17

```

```

Qy      186 DKLIDMAVSVKRHSRALGRQVVDMGIAVQTESALNVTTVL--APALSRPSVQGAVD 24
          | : :: | : : : | : : |
Db      174 -----AFVRANTPPTAAADIDLAVKAALGTIILNAATVSGIGYATATAAMIDLSD 22

```

```

QY      243 FGVST--AGGL-----VANAGFCDRLMLSVQSR-DQLRG-----GAFVLGAKD 28
      :| | | :      ::| | | : | | | | | | | | | | | | | | | |
Db      226 GASTDNAGVNFETAYPSSGVSGSTLSTTRTRDILLGTANNIDTFAVEGAGAILTVGD 28

```

Qy 282 KEPRAALSEETDILDAYKAIKSASYSGAALNACKRMAGLPIDVATDOLKAVRSVLSATSL 34  
| :: : | | : | | : : : : : :  
Db 286 TISGGAAGTMTNNVDA-----AAVTA-----IPTGVTISGTFMTNVTSGAAT 32

```

0y      342 TRIGLALAGYAVSKL-----QKATKNITDSATFAAQSLSNI-VGS 38
         | : | :: |           | :          | | :: | | 
000  AT- -GGCGTCTGCTATTATTTTCCTACCTGTCTCATTTAACTATTAATCCATCTTCA 39

```

QY 385 VGFAGMTTAGLATDPAVKKAESFIQDKVKTASTTSYADQTVKLAKTVKDMGSAIS 44  
|| :| || | | : | ::::|| :| | || :| :| :

QY 445 ST 446  
:1

## RESULT 14

RESULT 14	
AAW37490	
ID	AAW37490 standard; Protein; 1026 AA
XX	
AC	AAW37490;
XX	
DT	20-APR-1998 (first entry)

05 *Caulobacter crescentus*.

PN W09734000-A1.

PD 18-SEP-1997.  
xy

PF	10-MAR-1997;	97WO-CA00167.
XX		
PR	12-MAR-1996;	96US-0614377.
XX		

PA (UYBR-) UNIV BRITISH COLUMBIA.  
XX  
PI Bingle WH, Nomellini JF, Smit J,

DR WPI; 1997-470880/43  
DR N-PSDB; AAV01866.  
XX

PT S-layer protein - expressed as fusion proteins containing antigenic peptides in *Caulobacter*, useful as live vaccines

XX The present sequence represents  
CC example of the present invention  
CC

CC the DNA construct above additionally containing a sequence encoding a  
CC amino acids (aa) of Caulobacter S-layer protein. Caulobacter containing  
CC upstream of DNA encoding a C-terminal region of at least the last 82  
CC which contains at least one restriction site for insertion of DNA

CC heterologous polypeptide, are particularly useful in live vaccines  
CC (where the heterologous polypeptide is an antigen). They can also be  
CC used for production of e.g. ligands, enzymes or other proteins, e.g.  
CC metalloproteins to remove haemmetals from water or sewage or

CC xylanase or cellulase for use in wood pulping. All known *Caulobacter* CC strains are harmless, and stable in outdoor environments, including CC water (so suitable for vaccinating fish) or soil. They are well suited

CC an ideal system for presentation of antigens, at high level.  
XX  
SQ Sequence 1026 AA;

Query Match	5.5%;	Score 132;	DB 18;	Length 1026;
Best Local Similarity	21.0%;	Pred. No. 0.063;		
Matches 101;	Conservative 75;	Mismatches 166;	Indels 140;	Gaps 23;

[illegible]







13	135	5.6	973	2	C85693	probable membrane
14	134	5.6	2535	2	AC0304	probable hemolysin
15	132.5	5.5	654	2	C87587	hypothetical prote
16	132.5	5.5	1104	2	S59310	probable membrane
17	131.5	5.5	1214	2	G97419	streptococcal hema
18	131.5	5.5	1248	2	AH2637	conserved hypotet
19	131	5.4	409	2	S70348	sapc protein - Sal
20	130	5.4	1026	2	A48995	paracrystalline su
21	130	5.4	1073	2	C87374	S-layer protein Rs
22	129	5.4	436	2	G81655	conserved hypotet
23	129	5.4	637	2	C87322	methy]-accepting c
24	128.5	5.3	2388	2	J50271	beta spectrin, bet
25	127	5.3	397	2	I39579	necB protein - Alic
26	127	5.3	1010	2	AH2553	hypothetical prote
27	126.5	5.3	1131	2	T41144	hypothetical sear
28	126.5	5.3	1630	2	A53577	ascites staloglyco
29	126.5	5.3	2271	2	F90073	hypothetical prote
30	125.5	5.2	902	2	H87323	hypothetical prote
31	125.5	5.2	1225	2	F75605	minor tail protein
32	125	5.2	409	2	S70215	slpC protein - Sal
33	125	5.2	409	2	AH0850	pathogenicity isla
34	125	5.2	1329	2	T29074	hypothetical prote
35	124	5.2	1128	1	T08322	plasmid replicatio
36	124	5.2	2481	2	D90011	fatB protein [impo
37	123.5	5.1	860	2	T14650	tail fiber protein
38	123.5	5.1	978	2	T14968	phage lambda-relat
39	123.5	5.1	1147	2	T35781	hypothetical prote
40	123.5	5.1	2232	2	T34434	hypothetical prote
41	123	5.1	544	2	JC6063	chaperonin groEL -
42	123	5.1	544	2	B83720	class I heat-shock
43	122.5	5.1	789	2	E84236	Hcr6 transducer [i
44	122	5.1	6713	2	B89921	hypothetical prote
45	121.5	5.0	518	2	A55864	flagellin - Woline

# ALIGNMENTS

**RESULT 1**  
conserved hypothetical protein Atu1348 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C/Accession: AD2742  
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Kitejima, J.P.; Okura, V.K.; Almeida Jr., N.F.; Zhou, Y.; Bovee Sr., D.; Chapman, P.; Clendenning, J.; Decherage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kucyavin, T.; Levy, R.; Li, M.; McClelland, E.; Palmeri, A.; Raymond, C.; Rouse, G.; Steenphilmachak, C.; Wu, Z.; Gordon, D.; Eisen, J.A.; Paulsen, I.; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B.; Ilaeo, L.; Kim, S.; Hendrick, C.; Zhao, Z.; Dolan, M.; Tingey, S.V.; Tomb, J.; Gordon, M.P.; Olson, M.V.; Neeter, E.W.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AD2742  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-2155 <RUR>  
A/Cross-references: GB:AE008688; PION:AAI42354.1; PUD:g17739760; GSPDB:GN00186  
A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Gene: Atu1348  
A/Map position: circular chromosome

Query Match 6.2%; Score 149.5; DB 2; Length 2155;  
Best Local Similarity 20.7%; Pred. No. 0.35;

Matches 96; Conservative 65; Mismatches 176; Indels 127; Gaps 16;

QY	69	DALAHNEQIDALVETRANRLYEEGPATIAETFAKEKFDRLATTASAF-----ENTP	124
DB	609	DAFTNSHAKIDVYLAERSNALF-----GALSASQDRFDALASNLAITGSVGTAA	659
QY	125	FMAASVLYQWQPAINK-----GDWLATPLKPLTPLISGALSAMQDVGTRQMDRANGDLH	179
DB	660	EHLAAMDERRAAAINSVVADVERRLTETLETRAAAITGAVSGLEDRISTDLESRTAA---	716
QY	180	YLSTSPDKLDMAVSKRHSRPAIDRQVDMGJNAVQTSALNVVRYVRLPALASR-----	234
DB	717	-----LHDVSGAESRIADTLD-----GRTAALSASISGVEERIALDMDRSLSTD	762
QY	235	-----PSVQAVPVFGVSTAGGLVAN-----AGFGDMLSVQSR--DQLNG	272
DB	763	MTFANVERLERSETLDNRTSALTGIVASAEKTIAGLSDRTATGDDVAGATRIATLIDG	822
QY	273	-----GAFVIGKDKKEPKAALSEETDMLDAYKAISASYSGA-----	309
DB	823	RTAALNVVSGAEER-----IADALDSRTMALDMTFSGAEKIAEALDTRTALUGEL	874
QY	310	ALNAGKRAGLPLDVAIDGKAVRSIVSATSILRIGLALAGYAGVKIQKMATKNITDS	369
DB	875	VASAEIRIAG-ALDSRTDSLKTIV-----VSAGEERIT-DVLDL	910
QY	370	ATKAASVQLSNLVGSGVGFAGMTTATGLATPAVKAESFIQDVKXSTASSTTSYVADQTV	429
DB	911	RTMALDMSFSGVEEKITDIIDORTA-----ALKSVVAGVEDRIAGALDSTRTALSG---	961
QY	430	KLAATVKDMSGAISSTGASLNSVTNNLRHSRPAEADIEEGGIS	473
DB	962	-IVSGAEERIAEALDSRTIALDMTISGVEERIAEAMDARASSIS	1004

**RESULT 2**  
C97523  
hypothetical protein AGR\_C\_2490 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C/Species: Agrobacterium tumefaciens  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C/Accession: C97523  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, B.S.; Cao, Y.; Akenazi, M.; Halling, C.; Mullin, L.; Hountel, K.;

Gordon, J.; Vaudin, M.; Iartchouk, O.; Epp, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Flanagan, C.; Crowell, C.; Gursion, J.; Lomo, C.; Sear, C.; Strub, G.; Cleto, C.; Slater, S.  
Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* C58.  
A/Reference number: A97359; MUID:21608551; PMID:11743194  
A/Accession: C97523  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-2155 <KUR>  
A/Cross-references: GB:AE007869; P1DN:AAK07140.1; PID:g15156408; GSPDB:GN00187  
A/Genes: AGR\_C\_2490  
A/Map position: circular chromosome

Query Match 6.2%; Score 149.5; DB 2; Length 2155;  
Best Local Similarity 20.7%; Pred. No. 0.35;  
Matches 96; Conservative 65; Mismatches 176; Indels 127; Gaps 16;  
QY 69 DAALAHNEQIDALVETRRANRLYSRGETPATIAETFAKAEKFDRLATTASSAF----ENRP 124  
DB 609 DAFNSHAKIDTVLAERSNALF-----GALSASQDRFDEALASRLATGSGVGT 659  
QY 125 FAASVQLQYQPAINK-----GDMLATPLKPLTPLISGALSGAMDQVGTMMDRAGDLH 179  
DB 660 EHLAAMLEDERAALINSVADVRLTETLETRAAAITGVSGISDRISDTLESRTAA--- 716  
QY 180 YLSTPDKLDNAVAVSRHSPALGRQVNDGIVNPFALUVRVTLVPLALASR----- 234  
DB 717 -----LHDVVGAESRIADTLD-----GRTAALSSAISGVENIADNDSTRTSLD 762  
QY 235 -----PSVQGVDFGVSTAGLVAN-----AGFGDRMLSVQSR--DQLRG 272  
DB 763 MTFANVEERLSETLNRTSALTGVASAEKTAGALDSRTATFGDVAAGETRIAEFLDG 822  
QY 273 -----GAFVLGMDKEPKAALSEETMLDAYKAIKSASVSGA----- 309  
DB 823 RTAALNAVVGAEER-----IADALDSRTMALDMFTSGAEKIEALDTRTAALGEL 874  
QY 310 ALNAGKRMAGLPDVAIDGLKAVRSIVSATSLTRIGIALAGVAGVAKLQRMATKNITDS 369  
DB 875 VASAEIRIAG--ALDSRTSLKTV-----VSAGAEIRIT--DVLDG 910  
QY 370 ATKAASQSLNLVSGVFAQMTTAGLATDPAVRAKESFIQDKVKSASTTSVYADQTV 429  
DB 911 RTMALDMSFGSVEEKITDLDGRTA-----ALKSNAVAGEDRIAGALDSRTAALSG--- 961  
QY 430 KILAKTVKMGSGEALISSTGASLRSTVNRLRHNSAPEADIEGGIS 473  
DB 962 -IVSGAEIRIAEALDSRTIALDMTISGVETRIEAMDARASSLS 1004

RESULT 3  
AE3045  
Ice nucleation protein homolog [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)

C/Species: *Agrobacterium tumefaciens*  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C/Accession: AE3045  
R/Hood, D.W.; Setubal, J.C.; Kaul, R.; Monka, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Kitajima, J.P.; Okura, V.K.; Almeida Jr., N.F.; Zhou, Y.; Boyce Sr., D.; Chapman, P.; Clendinning, J.; Deatherage, G.; Gallet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClelland, E.; Palmeri, A.; Raymond, C.; Rouse, G.; Saenphimmachak, C.; Wu, Z.; Gordon, D.; Eisen, J.A.; Paulsen, I.; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B.; Liao, L.; Kim, S.; Hendrick, C.; Zhao, Z.; Dolan, M.; Tingey, S.V.; Tomo, J.; Gordon, M.P.; Olson, M.V.; Nester, E.W.  
A/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AE3045  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-989 <KUR>  
A/Cross-references: GB:AE008689; P1DN:AA44779.1; PID:g17742417; GSPDB:GN00187  
A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Genes: ina  
A/Map position: linear chromosome

Query Match 6.2%; Score 149; DB 2; Length 989;  
Best Local Similarity 22.4%; Pred. No. 0.13;  
Matches 127; Conservative 77; Mismatches 228; Indels 136; Gaps 25;  
QY 7 AQQPPGVANESFRRTASDASSVSVSTTSCRDQAI--TDVLRH--HVFALHNSVY 62  
DB 98 ASQIDAMSTDOIKALNASSQVAGLSSQVATISDELAFTTDEIKISANAIAGLSAAAI 157  
QY 63 GSPDERPAALANEGQIDALVETRRANRLYSRGETPATIAETFAKAEKFDRLATTASSAF----DRLATTA 116  
DB 158 AGLSTDAALALTKSQIAMSSTQFRNALTSGLATTSADDEVKAIKNTIAGLDVTKLSTON 217  
QY 117 SSAFENTPFAASVQLQYQPAINKGDMLATPLKPLTPLISGALSGAMDQVGTMMDRAG 176  
DB 218 IAAISKAQVSAALSTTQFAMSTDC-----IKALTSQVAGLSA--QVATLSSD---- 264  
QY 177 DLHVLSTSPDKLDNAVAVSRHSPALGRQVNDGIVNPFALUVRVTLVPLALASR----- 234  
DB 265 ELALFSTDEIKAIKANNVA--GLSAAALALTTNAAALTKTQIAGLSTQNALTSANLA 323  
QY 216 TFSALNVVTRVLAIPALASREVSQGVDFGVSTAGLVAN-----NAGFGDRMLSVQSRDL 270  
DB 324 TFSA--DEIKAIKISTKLA-----GLDVTKLSTQNALTTQQAASLSTQFAMSTDOI 375  
QY 271 RG--GAFVLGMDKEPKAALSEET--TMLDAYKAIKSASVGAALNA----- 313  
DB 376 KALTSQVAGLSSQVATISDELAFTTDEIKASFSANNAVAGLSTAAALATTONATALT 435  
QY 314 GKRMAGL--PLDVAIDGLKAVRSIVSATSLTRIGIALAGVAGVAKLQRMATKNITDS 370  
DB 436 KTOIAGLSTQNALTSGLIATFSADDEVKAIK--ALAG-----LDVTKLSTGNVA--AL 487





T12796  
 Probable transglycosylase - *Bacillus subtilis* phage SPB2  
 CDate: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #ext\_change 15-Oct-1999  
 CAccession: T12796; A69911  
 R:Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.  
 submitted to the EMBL Data Library, August 1997  
 A:Description: The complete nucleotide sequence of the *Bacillus subtilis* SPB2C prophage.  
 A:Reference number: Z17583  
 A:Accession: T12796  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2285 (AAZ)  
 A:Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025510; PIDN:AC13005.1  
 V:Kunt, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azorodo, V.; Bertero, M.G.; Bostleers, P.; Bolotin, A.; Borchert, S.; Borlas, R.; Boursier, L.; Brans, A.; Braun, M.; Bridgell, S.C.; Bron, S.; Brullier, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.-J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E  
 Nature 380, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Goldgity, E.J.; Grand, G.; Gutsapp, G.; Guy, B.-J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itay, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klsert-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapdash, A.; Lelidnols, S.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue, C.; Medigue, C.; Medina, N.; Molledo, R.P.; Mzuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogizawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle, D.; Porwoll, S.; Prescott, A.M.; Pressanc, E.; Pujic, P.; Puterle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Roche, R.; Roche, B.; Rose, M.; Sadele, Y.; Sato, T.; Scanlon, E.  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro, P.; Shih, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tarpetra, P.; Tononi, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzneger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zunstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MJD:98044033; PMID:9384377  
 A:Accession: A69911  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2285 (KUN)  
 A:Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CB14053.1; PID:el183582; PID:g2634555  
 A:Experimental source: strain 168  
 C:Genetics:



```

Db      855 ISATTIASLATQIALVALNVRISIAALSTVOIVA-----LTTAQVEALTTVOVGALST 906
Qy      163 MDQVGTMMDDRARGDLHTYSTSPDKLIDMANVYKRHSFALGRQVY-----208
Db      907 -----QLAVLST-----DIDATFSTRMALGSSNIGLSKDTVASLTTA 946
Qy      209 DKMG-----IIVOFFSALNVRITVLAPALASRP-----SYVGAVDEGVSTA 248
Db      947 QIGALSMAGISGLSTGQIALHGDQINVLNTQTIALHTSQQVAEVDSDITALSTGQIAA 1006
Qy      249 GGLVANAGFGDMHLSVOSRDQLRGAFVJGMRKDEPKAALSEE-----TWLDIAVPAIK 302
Db      1007 LSAAGGAAGLTTQIIAALSTDYV---GAMTSGQ-----IAALSKQIAALGDIATFSTGD 1056
Qy      303 SASVSGAALNAGRMAGPLVDVATDGLKAVRSIVSATSLT KNGALAGV-----352
Db      1060 IAAALSNVA-----VAGLSRDFVASLTTNQIALMSAGISGLGTQIAGLSTSEQVNVLTN 1116
Qy      353 AGVSKRL---QKMATNITDTSATTAVALSOLSNLVGSYGFAGMTTGLATDP---AVKKAE 406
Db      1114 AQISALTSRQVAALEVIDIASLSA-AQIA-A-ATGAAGV-----AGLTTDDIAALSTSQVE 1166
Qy      407 SFIDQKKTASTASTTSVVADQVRL-----ARTVKDMSEALSSSTGA 448
Db      1166 APTSQAQIALNKSQIAALSNADLAI FTTAEMALIGSALSGLPASTIASLTTQIALGA 1226
Qy      449 S 449
Db      1226 A 1226

```

Query Match 5.7%; Score 138; DB 2; Length 1731;  
Best Local Similarity 21.4%; Pred. No. 1.4;  
Matches 116; Conservative 76; Mismatches 197; Indels 152; Gaps 23;

QY 1 MHUNSOAQPVGAVNESFRTASDASVLASSSVSVSTTSQRDQALITDYLKHVFAAHRSVI 60  
Db 746 MHIAELGTE---LAKTTKDIAISSAIAIGLSA-----EALSLTTAQIAALNTQS 795  
QY 61 VIGSPDERDALAHNEQIDALVETRNRL-----YSEGETPATIAETFA- 104  
Db 796 IVALSTAQIAALT-TAQVEALSTQVNAALTSKQIAALSTDDIAFTSTKDIALINSDALAG 854  
QY 105 -KAEKTRLATATSSAFENTPPAAASYLDQWQPAINKGMWLAT-PLKPLTPLISGALSQA 162  
Db 853 LSAETIASLATQIIMAINVRSIALSTVOIVA-----LTTAQVEALITVQVGLSST 906  
QY 163 MDQVGTQWMDRARGDLHYLSTSPDKLHDMAVSVKRHSPALGRQV----- 208  
Db 907 -----QLAVLST-----DIAFTSTKDMAALSSAIGLSKDTVASLTTA 946  
QY 209 DMG-----IAVQTFSAIUVVRYVLAPALASRP-----SVQGVDFGVSTA 248  
Db 947 QIGALSMAGISGLSTQJIALTGQDLVLTNTQJIALTSKQVAAFVSDITIALSTQIAA 1006  
QY 249 GGLVANAAGFGDRMLASVQSRDQLRGGAFLVGRKDKERKALSEF-----TMDLDVYRAIK 302  
Db 1007 LSMAAGALTTDQIIMALSTIDV--GANTSGQ-----IMLSAKQIIMALGTDIAFTSTGD 1059  
QY 303 SAYSAGAILNAGKRNAGPLDVATPDGIAKNSIVSATSLTRKIGLALAGV----- 352  
Db 1060 IMAISSNA-----VAGLSRDTVASLTTAQIAALSSAGISGLGTQJIALGTSEQVNVITN 1113  
QY 353 AGVSKL--QKMATKNITDSATRKAAVSQLSNLVSGVGFAGWTTAGLATDP---AVKKA 406  
Db 1114 AQISALTSKQVAALEVTDIASLSA-AQIA-ALGAAV-----AGLTTDQIAALSISQVE 1165  
QY 407 SFIQDKVASTASSSTTSYVADQTVL-----AKTVKMGSEALISSTGA 448  
Db 1166 ALTSAQIAALNSKQIAALSDADLAFTTANMAAIGSGAISGLPASTIASLTTAQIAALGA 1225  
QY 449 S 449  
Db 1226 A 1226

RESULT 12  
B90835  
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7,  
substrain RIMD 0509952)  
C/Species: Escherichia coli  
C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C/Accession: B90835  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.;  
Han, C.G.; Ohnubo, E.; Nakayama, K.; Murata, T.; Tanaka, T.; Tida, T.;  
Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kihara,  
S.; Shibata, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7  
and genomic comparison with a laboratory strain K-12.  
A;Reference number: A99629; PMID:21156231; PMID:11258796  
A;Accession: B90835  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-971 (HAV)  
A;Cross-references: GB:BA000007; PIRN:BA835073.1; PIR:q13361114; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:  
A;Gene: ECs1650

Query Match 5.6%; Score 135; DB 2; Length 971;  
Best Local Similarity 22.0%; Pred. No. 0.96;  
Matches 116; Conservative 57; Mismatches 238; Indels 116; Gaps 20;

QY 3 INQSOAQPVGAVNESFRTASDASVLASSSVSVSTTSQRDQALITDYLKHVFAAHRSVI 62  
Db 110 VEEVARNASAVQNTAAKKSASDASTSAREALT----- 143  
QY 63 GSPDERDALAHNEQIDALVETRNRLVSEG--ETPATIAETPAKENTDRLATTSASF 120  
Db 144 HATDAADSAALASTSACQAAASQASASSAGTASTKATEASKAAAESSKAAATSAQA 203  
QY 121 ENTPPAAASYLDQWQPAINKGMWLAT-PLKPLTPLISGALSQANDQVGTMDRARGDLHY 180  
Db 204 ARTSETNAAVSQ--QSA-----ATSASTATKASEAASARBASAKKAS-SETSA 253  
QY 181 LSTSPDKLHDMAV--SVK-----RHSPLAGQVQVMDGIAVQTFAL--NWVRYVL 227  
Db 254 ASSASASASATPAAGNSAKAAKSETNAKSEPTAAEQSAAAGSTMAAASASTA 313  
QY 228 APALASRPVQGVDFGVSTAGLVANAGFGDRMLASVQSRDQLRGGAFLVGRKDKERKAA 287  
Db 314 QGASASATPAAGNSAESASASTATTKAGATEQASA-----AASASAAKSETNAK 366  
QY 288 LSEETDMLDYRAIKSASVSGAALNAGKRNAGPLDVATPDGIAKNSIVSATSLTRKIGLA 347  
Db 367 ASETV--AESSTKTAASASASASSASS--ASASKDEATRQASAKS--SATASTATE 420  
QY 348 LAGVAGVSKLQ-----KMATKNITD-----SATKAAVSQLSNLVSGV 387  
Db 421 AAGSATAAQSSTAESATRAETAAKAEEDIASAVALEDSITTKGIVQLSATNS--- 477  
QY 388 FAGWTTAGLATDP-AVKKAESFIQDKVASTTSYVADQTVYKLAFTVKMGSEALISST 446  
Db 478 -----TSESLAATPKAVKAAVELANGK-----YTAQDATTAQKGIQVLS-NATNST 522  
QY 447 GASLRSTVNNLR-----HRSAPEDIEBGGI-----SAFSRSET 480  
Db 523 SMDLAATPKAVKAAVDIANGKRYTAQDATTAQKGIQVLSATNSASSET 569

RESULT 13  
C85693  
probable membrane protein of prophage CP-933X Z1918 [imported] - Escherichia  
coli (strain O157:H7, substrain EDL933)







OM protein - protein search, using sw model

Run on: January 6, 2004, 09:37:23 ; Search time 18 Seconds

(without alignments)  
1272.334 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407  
Sequence: 1 MHINQSQPPGVAMSFRT.....EECGISAFNSRSETPQLRL 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.1\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	138	5.7	1120	1 STFR_ECOLI	P76072 escherichia
2	137	5.7	540	1 CH60_LACHE	068324 lactobacilli
3	130	5.4	500	1 FLDB_SALAE	P52615 salmonella
4	130	5.4	1025	1 SLAP_CAVCR	P35828 cauldobacter
5	129	5.4	436	1 Y868_CHLMU	Q9pj1 chlamydia m
6	128.5	5.3	2388	1 SPCP_RAT	Q9qmr8 rattus norv
7	128	5.3	546	1 CH60_LACAC	Q93607 lactobacilli
8	127	5.3	397	1 NCCB_ALCX	Q44565 alcaligenes
9	123	5.1	344	1 CH60_BACHD	Q50305 bacillus ha
10	121.5	5.0	1068	1 H1PR_MOUSE	Q93ky5 mus musculu
11	120.5	5.0	1140	1 YW96_YEAST	Q04893 saccharomyc
12	120	5.0	1306	1 MSB2_YEAST	P32334 saccharomyc
13	119.5	5.0	1608	1 HLYA_SERMA	P15320 serritlia ma
14	119	4.9	657	1 H57F_CABEL	P11141 caenorhabdi
15	118.5	4.9	760	1 MCFE_MYCLE	Q05364 mycobacteri
16	118.5	4.9	880	1 LYTD_BACSU	P39848 bacillus su

17	117	4.9	583	1 YP65_MYCTU	Q50733 mycobacteri
18	117	4.9	682	1 GR78_YEAST	P16474 saccharomyc
19	116.5	4.8	570	1 FLIF_RHOSH	Q53151 rhodobacter
20	116	4.8	763	1 HTR2_HAUNI	Q9np81 halobacteri
21	116	4.8	764	1 HTR2_HAUSA	P71410 halobacteri
22	116	4.8	1577	1 HLYA_PROMI	P16466 proteus mir
23	116	4.8	3591	1 FHBX_BORRE	P12255 bordetella
24	115.5	4.8	401	1 YOPB_YERPS	Q06114 yersinia ps
25	115.5	4.8	1065	1 SED4_YEAST	P25365 saccharomyc
26	115	4.8	679	1 GR78_KULJA	P22010 kluyveromyc
27	115	4.8	1239	1 V120_EBV	P03189 Epstein-Bar
28	113.5	4.7	571	1 FLAI_CAME	P56963 campylobact
29	113	4.7	2541	1 T1M1_TUMAN	Q9y490 homo sapien
30	112.5	4.7	505	1 FLDB_SALT	P52616 salmonella
31	112.5	4.7	2090	1 N214_HUMAN	P35658 homo sapien
32	112	4.7	439	1 Y579_CHLTR	O84583 chlamydia t
33	112	4.7	492	1 FLIC_SALMU	P06175 salmonella
34	112	4.7	1027	1 AF10_HUMAN	P55197 homo sapien
35	111.5	4.6	544	1 CH60_PROAC	Q9K2U4 propionibac
36	111.5	4.6	547	1 CH60_ALTHA	Q9K4U7 altomonas
37	111.5	4.6	553	1 MCP2_ECOLI	P07017 escherichia
38	111	4.6	881	1 PRV3_YEAST	P47033 saccharomyc
39	110.5	4.6	543	1 CH60_MYCGE	P47632 mycoplasma
40	110.5	4.6	1419	1 ALAI_CANAL	O13368 candida alb
41	110	4.6	545	1 HTP5_HAUNI	Q48318 halobacteri
42	110	4.6	549	1 CH60_COLMA	Q93618 colwellia m
43	110	4.6	814	1 SLAI_BACAN	P49051 mus musculu
44	110	4.6	7389	1 BPAL_MOUSE	Q91266 mus musculu
45	109.5	4.5	901	1 SOK1_YEAST	P40317 saccharomyc

ALIGNMENTS

RESULT 1	
ID STFR_ECOLI	STANDARD; PRT; 1120 AA.
AC P76072; P77560;	
DT 16-OCT-2001 (Rel. 40, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	
DE Side tail fiber protein homolog from lambdaoid prophage Rec.	
OS STFR OR B1372.	
OS Escherichia coli.	
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
OC Enterobacteriaceae; Escherichia.	
OX NCBI_TaxID=562;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=K12 / MG1655;	
RX MEDLINE=97426617; PubMed=9278503;	
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,	
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,	
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,	
RA Mau B., Shao Y.;	
RT "The complete genome sequence of Escherichia coli K-12.";	
RL Science 277:1453-1474(1997).	



[illegible]

QY 56 AHRFVSIVSGPDEDAALANRQIDALVETFRANRLYSGETPATLTAETFAAKETKPRLATTT 115  
D 74 AH--TTEGALNEI--NNNL--RRVRELAVOSANSTNSQSDLSIQAEITQRHLEIDRY--S 126  
QY 116 ASSAFENTPFAASVQVYQWPAIKNGKMDLAPLPPLISGA-----LSGAMDQVGTWMM 171  
D 127 GQRQFNQKVLQADNTLRTQYGANDGETIDIDLQINSQTLGLDSLVQVQKXDVSADVI 186  
QY 172 DRARGLHYLSTSPDKLHDMAVSVKRHSPLAQOV--DMGIADVTFSAIWNVETVLAP 229  
D 187 STSYSGTQALTP-----TATDIK---AALGPVTGTGDTLTAASFCKGYKATVSGY 237  
QY 230 ALASRVSQVGNDFGVTAGLVANMGFGDRMLSVSQRQLGGAFLVGRKQKEKVALS 289  
D 238 TIDAGTANRQKGYETVYDSTAGVS---FQ---ATPKSTVGTDAATKQVNAVA-- 288  
QY 290 EETDWLDVYKAIKSASVSGALANRGMAGLPDLVATDGLKAVRSIVSATSLTKRGLALA 349  
D 289 -----DA--ATKRLAQDQGVSSADSA-----TIVKSYIDRKGITIE 325  
QY 330 GGYA-----GVSKIQKATKNTIDSAKRAVSQLSNLVSGVFGAGMTT 393  
D 326 GGVALKAGDRYAADYDEATGAIK-ARTSTSYAADGTTKTAANLQGVDEKTEV---T1 381  
QY 394 AGIADDPVAKAESF-IQDKKSTASTSTSYVAQ--TVLAK-----TVKMGSEA 442  
D 382 DGTATVNAKKAAGIDFAQPELAEEAAATTEENLDKIDALAAQVADALRSIDGAVQNRNSA 441  
QY 443 ISTGASLRSTVNNLRHRSAPADIEEGC---ISAFSRSE 479  
D 442 ITNLG---NTVNNL---SEARSTIEDSDYATEVSNMSRAQ 475

RESULT 4  
SLAP CAUCR STANDARD; PRT; 1025 AA.  
ID SLAP CAUCR STANDARD; PRT; 1025 AA.  
AC P35828; 046015; 09RF12;  
DT 01-JUN-1994 (Rel. 20, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE S-layer protein (Paracrystalline surface layer protein).  
OS NSAA OR CCI007.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
OC Caulobacteraceae; Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=93007489; PubMed=1393820;  
RA Gilchrist A., Fisher J.A., Smit J.K.;  
RT "Nucleotide sequence analysis of the gene encoding the Caulobacter  
crescentus paracrystalline surface layer protein.";  
RL Can. J. Microbiol. 38:193-202(1992).  
RN [2]  
RP REVISIONS TO 376; 636 AND 842-843.

RL Submitted (OCT-1999) to the EMBL/GenBank/DBD databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J53001;  
RA Bingle W.H., Awram P.A., Nornellini J.F., Smit J.K.;  
RT "The secretion signal of C. crescentus S-layer protein is located in  
the C-terminal 82 amino acids of the molecule.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBD databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
Eisen J., Heidelberg J.F., Alley M.R., Ohta N., Maddock J.R.,  
Petrocke I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
Deboy R.T., Dodson R.J., Durkin A.S., Gaim M.L., Haft D.H.,  
RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
RN [5]  
RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=89008089; PubMed=3049545;  
RA Fisher J.A., Smit J.K., Agabian N.;  
RT "Transcriptional analysis of the major surface array gene of  
Caulobacter crescentus.";  
RL J. Bacteriol. 170:4706-4713(1988).  
RN [6]  
RP CHARACTERIZATION.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=98292737; PubMed=9620954;  
RA Awram P., Smit J.K.;  
RT "The Caulobacter crescentus paracrystalline S-layer protein is  
secreted by an ABC transporter (type I) secretion apparatus.";  
RL J. Bacteriol. 180:3062-3069(1998).  
CC -I- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A  
PHYSICAL BARRIER TO PARASITES AND LYTIC ENZYMES.  
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER  
(TYPE I) SECRETION APPARATUS.  
CC -I- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE  
SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A  
SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.  
CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
CC  
DR EMBL; AF062345; AAC3865.2; -.

DR EMBL; AF193063; AA19365.1; -.  
DR EMBL; AE005779; AA22991.1; ALT\_INIT.  
DR PIR; A48995; A48995.  
DR HSSP; P22629; 1SWC.  
DR TIGR; CC1007; -.  
DR InterPro; IPR001343; HemLysn\_Ca\_bind.  
DR Pfam; PF00353; hemolysincbind; 3.  
DR PRINTS; PR00313; CABINDGRPT.  
KW Cell wall; S-layer; Calcium-binding; Complete proteome.  
FT INIT MET 0  
SQ SEQUENCE 1025 AA; 98001 MW; ADY326E1363D8AC CRC64;  
  
Query Match 5.4%; Score 130; DB 1; Length 1023;  
Best Local Similarity 21.0%; Pred. No. 2.3;  
Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;  
  
QY 22 SDASLASSVSSTTSCHRDQAITDYLNHFARFVIGSGSDERDAALAHNQIDL 81  
DB 41 SDAAALNTLTLLNSTTVAALQT-----YQFFGVAFSAAG-----IDFL 80  
QY 82 VETRAM-----RLYSEGETPATIAETFAKAEK-----DRLATTSASEENTPEAAA--- 128  
DB 81 VDSTNTNDLNDAYSK-----FAGENREINFSINLATGAGC--ATAFAAYTG 128  
QY 129 -SVLDYVQPAINK--GDWLATPLKPLPLISGALSGAMDQVGTMMEDRARGLHYLSTSP 185  
DB 129 VSYAQTVATVADKIIGNAVAT-----AGGVDAVAVAFLSRQANIDYLT--- 172  
QY 186 DKLDHMAVSVKRHSPLAGROYVDMGIAVQTFSAALNVFTV---APALSPSYQGANV 242  
DB 173 -----AFVRANPTFTAAADIDLANAKAALIGTILMAATVSGIGYATITAMINDLSD 224  
QY 243 FGVSF--AGGI-----VANAGFGDRMLSVQS--RDQLRG-----GAFYLQKRD 281  
DB 225 GALSTDNAAGVNLFTAYPSGVSGLTLTGDTLTGRANNDFVAVAGEAQAATLVGD 284  
QY 282 KEPKAALSEETDMLDAYKAIKASYSGAALNAGKRAAGLPDVAATDGLAKVRSLSVATSL 341  
DB 285 TISGAGTDLNVWQA-----AAVTA-----LPTGVTISGIEETM-NVTSQAAI 326  
QY 342 TRNGIALAGYAGVSKL-----QKMATNIIIDSATKAAVSQLSNL-VGS 384  
DB 327 TLN---TSSGVTGLTALVNTNSGAQOTTAGAGQLATTTAAQAANNVAVDGAANVTVAS 383  
QY 385 VGVFAGMTTAGLATDPAVKAESFIODKVKSTASTTSYVADQTVKLAKTVKDKSGEALIS 444  
DB 384 TGVTSGTTTG-----ANSASGTVSVANSSTTTTGALIA-VTGGATVTVAGTGAANN 437  
QY 445 ST 446  
DB 438 TT 439

RESULT 5  
Y868\_CHLMU  
ID Y868\_CHLMU  
AC Q9PUG1;  
STANDARD; PRF; 436 AA.

DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein TC0868.  
GN TC0868.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Mopn / Nig9;  
RX MEDLINE=20150255; PubMed=10664935;  
RA Read T.D., Brunham R.C., Shen C., Gail S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Uitterlind T., Berry K., Baas S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwinn M., Nelson W., Debey R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
CC -1 SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0808/CTS79/TC0868  
CC FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
DR EMBL; AE002353; AA39664.1; -.  
DR PIR; G81655; G81655.  
DR TIGR; TC0868; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 436 AA; 43891 MW; 43E21C924CD7A48 CRC64;  
  
Query Match 5.4%; Score 129; DB 1; Length 436;  
Best Local Similarity 20.1%; Pred. No. 0.9;  
Matches 85; Conservative 67; Mismatches 170; Indels 100; Gaps 14;  
  
QY 66 DERDAAALNHEQIDLVETRAARNLVSEGETPATIAETFAKAEKDDL-----ATTASAF 120  
DB 30 DRNGAAAKTQOQEEASFEDLIQDSQGTGSKSKASTSQTSKGSEAKQKSSGTTTTTVA 89  
QY 121 ENTPEAASVLYQVMPAINKDWLATPLKPLPLISGAL-----SGAMDQVGTMMEDRAR 176  
DB 90 QASQATRAQAVKGAASSGFRNDDASLPSPNRRNVNGVLLKRNNGTLLALMGLINTLLAQ 149  
QY 177 DLHYLSTSPDKLHDMNAVSVKRHSPLAGROYV---DMGIAVQTFSAALNVFTVLAAPLAS 233  
DB 150 SAKSWSS-----SFOQNOAIIQNGVAVAPETGAIARTQAHQAATELQAKOSL 198  
QY 234 RPSVQGVADFGVSTAGLY--NAGFGDRMLSVQSROLRGAFVLMKHKPKKALSEET 292  
DB 199 ISGITNIVGRAVSVGGGISLASKSLG-----GLK-----SAFTNET 235

CC		are the predominant cell-type to express the gene. Found abundantly in Purkinje cells.
CC	-1- SIMILARITY:	BELONGS TO THE SPECTRIN FAMILY.
CC	-1- SIMILARITY:	Contains 2 calponin-homology (CH) domains.
CC	-1- SIMILARITY:	Contains 1 PH domain.
CC	-1- SIMILARITY:	Contains 17 spectrin repeats.
CC		-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by, and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/</a> announce/ or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC		EMBL; AB008551; BAA32699.1; -;
DR	EMBL; AB001347; BAA32473.1; -;	
DR	EMBL; AF225960; M425596.1; -;	
DR	PIR; JEO271; JEO271.	
DR	HSSP; Q01082; IBKR.	
DR	GO; GO:0016363; C:nuclear matrix; TAs.	
DR	GO; GO:0000809; C:spectrin; TAs.	
DR	GO; GO:0003779; F:actin binding activity; ISS.	
DR	GO; GO:0030506; F:ankyrin binding activity; ISS.	
DR	GO; GO:0005200; F:structural constituent of cytoskeleton; TAs.	
DR	GO; GO:0007158; P:regulation of calcium ion dependent exocytosis; TAs.	
DR	GO; GO:0016081; P:synaptic vesicle docking; TAs.	
DR	InterPro; IPR001589; Actbind actinin.	
DR	InterPro; IPR001715; Calponin-like.	
DR	InterPro; IPR001849; PH.	
DR	InterPro; IPR002017; Spectrin.	
DR	InterPro; IPR001605; Spectrin_PH.	
DR	Pfam; PF00307; CH; 2.	
DR	Pfam; PF00169; PH; 1.	
DR	Pfam; PF00435; spectrin; 17.	
DR	PRINTS; SM00683; SPECTRINH.	
DR	SMART; SM00233; CH; 2.	
DR	SMART; SM00233; PH; 1.	
DR	SMART; SM00150; SPEC; 16.	
DR	PROSITE; PS00019; ACTININ_1; 1.	
DR	PROSITE; PS00020; ACTININ_2; 1.	
DR	PROSITE; PS50021; CH; 2.	
KM	PROSITE; PS50003; PH DOMAIN; 1.	
KM	Cytoskeleton; Membran; Repeat; Actin-binding; Capping protein.	
FT	DOMAIN	1 278
FT	DOMAIN	57 161
FT	DOMAIN	176 278
FT	REPEAT	305 415
FT	REPEAT	425 529
FT	REPEAT	531 639
FT	REPEAT	641 745
FT	REPEAT	747 850
FT	REPEAT	852 956
FT	REPEAT	958 1063
FT	REPEAT	1065 1170
FT	REPEAT	1172 1276
FT	REPEAT	1278 1381
FT	REPEAT	SPECTRIN 10.



FT REPEAT 1383 1486 SPECTRIN 11.  
FT REPEAT 1488 1586 SPECTRIN 12.  
FT REPEAT 1588 1692 SPECTRIN 13.  
FT REPEAT 1694 1799 SPECTRIN 14.  
FT REPEAT 1801 1905 SPECTRIN 15.  
FT REPEAT 1907 2011 SPECTRIN 16.  
FT REPEAT 2013 2075 SPECTRIN 17.  
FT DOMAIN 2218 2328 PH.  
FT CONFLICT 326 328 VTL -> GTF (IN REF. 3).  
FT CONFLICT 543 543 L -> F (IN REF. 2).  
FT CONFLICT 608 608 D -> G (IN REF. 3).  
FT CONFLICT 887 887 L -> P (IN REF. 3).  
FT CONFLICT 908 908 V -> I (IN REF. 3).  
FT CONFLICT 948 948 D -> G (IN REF. 3).  
FT CONFLICT 1156 1157 EL -> GA (IN REF. 2).  
FT CONFLICT 1194 1194 F -> V (IN REF. 2 AND 3).  
FT CONFLICT 1555 1555 G -> R (IN REF. 3).  
FT CONFLICT 1769 1769 R -> W (IN REF. 3).  
SQ SEQUENCE 2388 AA; 271062 MW; 3EC8966AF065F19 CRC64;

Query Match 5.3%; Score 128.5; DB 1; Length 2386;  
Best Local Similarity 21.8%; Pred. No. 8.4;  
Matches 101; Conservative 51; Mismatches 168; Indels 143; Gaps 18;

QY 35 STTSCRDLOAITDYLRKHYFAAHFFSVI-----GSPDRDALA---LVETRAN 87  
DB 666 SAETGDLTGVLRLKHK--PALNGESGRLOPKLTLBOQQLVAEGHPGANOASTRAA 723  
QY 88 RLVSSEGETPATIAETFAKAEKFDRLATTASAFENPPFAAASVLOVQPAINKGMILATP 147  
DB 724 ELDAQWERLEALAE-----ERAKRLA-----QASLIVQFOADNDMEALVDA 766  
QY 148 LKPLPPLISGALSAMDQVGTAKMDRARGDLHYLSTSPDKLHDAAVSVKHSPPALGRQV 207  
DB 767 LR-----LVSSPEVG-HDEFTQALARQ-----HRALEEIRAAHP----- 801  
QY 208 VMGIAVQTFSAIINVRIVLAPALASRPVQAVDFQVSTAGLVANAAGGR----- 260  
DB 802 -----TIDALREQAALPPALSHTPPEVGRVPTLEOHYEELQARAGRALEAALA 853  
QY 261 ---MLSVQRDQLRGCAFVLRGKDKPEPKALSEETIMDLAKATKSAASYSGAALNAAGRM 317  
DB 854 FYTMLS-----EADACGLWEKE-----QMLN----- 876  
QY 318 AGLPIDVATDGLKAVNSIVSATSLTRNGALAGVAGVSKIQRMATNITDSATKAASVQ 377  
DB 877 -GLALPERLEDLELVQORFETLEPEENALA-----ARVAVSDIAEQLLKASPPG 925  
QY 378 LSNLVGS-----VGVFAGMTTAGLATDPAVK-----KAESFIODVKSSTAS- 419  
DB 926 KDRITGTEQQLNQRMQFRSLADGKRAALTSLSIQVYHLECTETQAMKREKTVISTQ 985  
QY 420 -----TTSVADQTVKLAKTVKDSGEAISSTGASLSTNVNL 457  
DB 986 DLANDLAGVIALQR-KIAGTERDL--EALSARVGEIQLQANAL 1025

RESULT 7  
CH60\_LACAC  
ID CH60\_LACAC STANDARD; PRT; 546 AA.  
AC 093607;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 60 kDa chaperonin (Protein cpn60) (groEL protein).  
GN GROEL OR GROEL.  
OS Lactobacillus acidophilus.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1579;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CRL 639;  
RA Lorcea G.L., Font de Valdez G.;  
RT "Characterization of the Lactobacillus acidophilus CRL 639 groESL operon."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -I- FUNCTION: Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity).  
CC -I- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of 7 subunits (By similarity).  
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -I- SIMILARITY: Belongs to the chaperonin (HSP60) family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
DR EMBL; AF300645; AA097218.1; -.  
DR HAMAP; MF\_00600; -. 1.  
DR InterPro; IPR001844; Chaperin cpn60.  
DR InterPro; IPR002423; Cpn60/TCF-1.  
DR Pfam; PF00118; Cpn60\_TCF1; 1.  
DR PRINTS; PR00298; CHAPERONIN60.  
DR PRINTS; PR00304; TCOMPLEXTCP1.  
DR PROSITE; PS00296; CHAPERONINS\_CPN60; 1.  
KW Chaperone; ATP-binding.  
SQ SEQUENCE 546 AA; 58222 MW; CFAF52B2217513F4 CRC64;

Query Match 5.3%; Score 128; DB 1; Length 546;  
Best Local Similarity 21.2%; Pred. No. 1.4;  
Matches 111; Conservative 71; Mismatches 196; Indels 146; Gaps 26;

QY 21 ASDASLASSSVSSTSCRDLOAITDYLRKHYFAAHFFSVI-----GSPDRDALA--- 73  
DB 2 AKNDIKFAENARSL-----LKGVDKLDADVTTTIGKGNVVLQSGNPDTINDGVTIAK 57  
QY 74 -----HNEQIDALVETRA-----NRLVSEGETPATIAETFAKAEKFDRLATTAS-----S 118

Db 58 STELKHVENNKAIVAEAAQKINDIAGXGTTATVITQALAREGKANTAGANFVGIR 117  
QY 119 AFENPFPAASVLOVQWOPAINKGEWLA--TPKLPFLPISGALGMDVQ----- 167  
Db 118 GIEKATKAADDELHKI SHKVESKQIANYAAVSSASKEVGLIDAMENKVGHEVITTED 177  
QY 168 -----TRKQDRANGDL-HYLSTSPDKLH-----DANAVSKRHSPLGR 205  
Db 178 SRCINTELSTVEGHEQFDRGYLSQYMTDNDKREADLNPYILITDKRTSINQDILPLIL-Q 236  
QY 206 QVVDGIAVQTESALNVVRYTLAPALAS--RPSVQGVDFGSTRAGSLVANKGFGDRLS 263  
Db 237 EIVQGG-----KSLILIDDDVTGEALPTLVANKIRGT--FNVA-----YAAPGFQDRK-K 284  
QY 264 VQSRD--QLRGAFAV---LQMKQKEPPALSEETDMLDAIKASVASGAILNAGKRNA 318  
Db 285 AQLEDIALTGTGVITVDLGFELMDTK----- 311  
QY 319 GLPLDVAITDGLKAVNSLVATSITKNGLALAGVAGVSK-----LQKMATKNTDLSA 370  
Db 312 -----IDQLGQARRV---TYTKDSTIYDQ-AGSKAIKEREDSIRKQIEESTSDFD 359  
QY 371 TRAAVSQSLNVGSGVGFAGWTTAGLATDPAVKAESEFIQDKVAST-ASSTSYVA---D 426  
Db 360 KKKLOERLAEITGGAVI---HWGAATETELKERKRIEDIASTRAAVDEGIVAGGCT 415  
QY 427 QTVKLAKTVKDMSE-ALSTGASLSTVNNLRHSAPAEADIE 469  
Db 416 ALVDVEKAIKDLKGETSDEQGINI-----VLRLASAPRQIAE 454  
RESULT 8  
ID NCCB\_ALICXX STANDARD: PRT; 397 AA.  
AC 044585;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DE Nickel-cobalt-cadmium resistance protein nccb.  
GN NCCB.  
OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Achromobacter.  
OX NCCB\_TaxID=515;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=31A;  
RX MEDLINE=95050278; PubMed=7961470;  
RA Schmidt T., Schlegel H.G.;  
RT "Combined nickel-cobalt-cadmium resistance encoded by the ncc locus  
of Alcaligenes xylosoxydans 31A.";  
RL J. Bacteriol. 176:7045-7054(1994).  
CC -I- FUNCTION: COMPONENT OF THE NCC CATION-EFFLUX SYSTEM THAT CONFERS  
CC RESISTANCE TO NICKEL, COBALT AND CADMIUM.  
CC -I- SIMILARITY: HIGH, TO A. EUTROPHUS CNB AND SOME, TO ALCALIGENES  
CC CZCB.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L31363; AAA65105.1; -.  
DR PIR; I39579; I39579.  
DR InterPro; IPR006143; HlyD.  
DR InterPro; IPR005693; Membrane\_fus2.  
DR Pfam; PF00529; HlyD; 1.  
DR TIGRFAMs; TIGR00999; Bae102; 1.  
KW Plasmid; Nickel; Cobalt; Cadmium; Transport.  
SQ SEQUENCE 397 AA; 39933 MW; C6DE57CE314996B8 CRC64;  
Query Match 5.38; Score 127; DB 1; Length 397;  
Best Local Similarity 22.88; Pred. No. 1;  
Matches 89; Conservative 59; Mismatches 140; Indels 102; Gaps 19;  
QY 149 KPLPFLPISGALSGAMDQVGRKMDRANG-----DLHYLSTSPDKLHDMAVSVKRHS 201  
Db 9 QPSNPMIAG-VAAAAALVG---FGAARGLSPSGAEVKLAAPER---AAASAPALP 60  
QY 202 A-----LGRQVYDMGIAVQTESALNVVRYTLAPALASRPSVQGVDFGVSTAGLVANAGF 257  
Db 61 AAVRIPGEYLAANINAVEPVSAGVGSVLLAP--ASVAAPVGSBAVVASRAAGV----- 113  
QY 258 GDRMLSVQSR--DQLRGCAFVLMKDKPEPKALSEETDMLDA-----YKAIKSNIS 307  
Db 114 -----LRIQRLGDPAVFRAGD-VIALVDSEPAAMAAERKVAQARADLARKTYERESSLPQ 168  
QY 308 GAALNAKRRVAGPLDVAITDGLKAVNSLVATSITKNG--LALAGVAGVSKLQKMATKN 365  
Db 169 GVTTPROMESAKRILDAVQAEVQRAATVQAQKAVSSDGRSVAVVSTIAG-----RITQS 223  
QY 366 ITDSATYAAVSQLSNLVGSGVGFAGWTTAGLATDPAVKAESEFIQDKVSTASTSYVA 425  
Db 224 VTLGAVVAPOAELEFRVAGSAGV-----QVEAVY-----TAADTSRIA 261  
QY 426 --DQTVKLAK-----TYKDMSGEALISST-----GASLR--STVN 455  
Db 262 GSDATIVLAGAPLAGRVQAVTPVSGSAPAAATVVTPEVDMNSGLIVGEGVQVRLTKAA 321  
QY 456 NLRRHSAPAEADIE--EGGISAFSRESETPQ 483  
Db 322 DANMASVPEDAVQNLDRGDRVVFRTQGGFR 351  
RESULT 9  
ID CH60\_BACHD STANDARD: PRT; 544 AA.  
AC 050305; Q9KFC3;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)

FT	CONFLICT	500	500	A -> VP (IN REF. 1).
FT	CONFLICT	512	512	MISSING (IN REF. 1).
SEQ	SEQUENCE	544 AA;	57403 MW;	CT69EDB1CS03C971 CR664;
	Query Match	5.18;	Score 123;	DB 1; Length 544;
	Best Local Similarity	18.98;	Pred. No. 2.7;	
	Matches 104;	Conservative 75;	Mismatches 192;	Indels 180; Gaps 23
QY	21	ASDASLASSSVASVTSQCDQAIPTYLK-----HHFAAHFVSGSP-----	65	
DB	2	AKNDIKRSEDARSH---LRGDKLDANKVTLQPKGNVLEKFF---GSPILINDQVT	54	
QY	66	-----DERDALAHNEQIDALVETRANRLYSEGETPATIETFAKEKEDRLATTASSAF	120	
DB	55	IAKEIELEDAFENKAGLVAEVAASHKIDIAAGGTTATYV---LAQAMIREGLRVTSGA-	110	
QY	121	ENTFFPAASVLDVQNPALINKGMATPLKRLPLSL-GALSGMQRVQ--TRQDPRARG	176	
DB	111	-HPWVIRKGEIAQVAVVEELSKISKPIEGKDSIAQVAAISSADEGKTIIAENAEVRGN	169	
QY	177	D-----LHYLSTSPDKLH-----DAAVASVK	197	
DB	170	DQVITTFESKSGTELEVVEGQDFDRGVASPIYMTDSDDKAEVLDNPVYLTIDKTSINIQ	229	
QY	198	RHSPALGROYDMGIAQVTFSSALNVETVLAPLASRPSVQANDFGVSTAGLVAN---	254	
DB	230	EVLTVL-EQVVOGQKPI-----LIIAEDVEGEALAT-----LVNKLRL	266	
QY	255	-----AAGFGRLTVSQRQRLGAGAFVLMKCKEPRKALSEETDMLVYRAIKSA	304	
DB	267	GTENAAVVAKPFQFGR-----RKAMLEDD-----	290	
QY	305	SVSGALNAGKRWAGLPDVAITDGLKAVRSLSVTSILTKNGLALAG-----YAGVS	356	
DB	291	-----ALITGGEVITFDGLDKSNITQL-GRASVYVTKENTTIVEGAGESDKTAAARN	345	
QY	357	KIQGKRAITRTSATKRAVQSQSLVNSVGQVAGHTTAGLTADPAKKAESFIQKVKST	416	
DB	346	QIKQIQEETTSDEFEKLERLAKIAGAAVVI-----KVGAATEEMKERLRIEDALNST	401	
QY	417	ASSTTSV-----ADQVTKLAKTVKDNKGSAISSTGA-----SLRSTVNNLRHSAPLEADI	467	
DB	402	RAVEEBEIVAGGTRALNNVIKAVSSIAEBEDGATQWIVLRALBEPVRLAHNAAGLESV	461	
QY	468	-----EEGG	471	
DB	462	IVERLKKKEAG	472	
RESULT 10				
HIPR_MOUSE	ID	HIPR_MOUSE	STANDARD:	PRT; 1068 AA.
AC	Q9URK5;			
DT	16-OCT-2001 (Rel. 40;			Created)
DT	16-OCT-2001 (Rel. 40;			Last sequence update)
DT	16-OCT-2001 (Rel. 40;			Last annotation update)
DE	Huntingtin interacting protein 1 related (H1pi-related).			

[illegible]



FT DOMAIN 698 816 7 X 17 AA TANDEM REPEATS.  
FT REPEAT 698 714 1.  
FT REPEAT 715 731 2.  
FT REPEAT 732 748 3.  
FT REPEAT 749 765 4.  
FT REPEAT 766 782 5.  
FT REPEAT 783 799 6.  
FT REPEAT 800 816 7.  
SQ SEQUENCE 1306 AA; 133114 MW; 67D5D984D5CA46D CRC64;

Query Match 5.0%; Score 120; DB 1; Length 1306;  
Best Local Similarity 20.2%; Pred. No. 12; Indels 148; Gaps 24;  
Matches 117; Conservative 79; Mismatches 235; Indels 148; Gaps 24;

QY 14 AMESFRASDASLAS-----SSVRSVSTTSCRDQAITDYLKHHVFAAHREFSVIGSPDE 69  
DB 142 ASSVMAASDVSTASVPIDTSNANSIPPTTTSNIETTT-----SAPLUSD 185  
QY 70 AALANEOIDALVETRANRLYSEGETPATIETAKAEKFDRLATASSAFENT--PEPAA 128  
DB 186 TPLISTSTMSA-----ADVFSANPISALITTDSESEFDSTAGALPVSADFESS 240  
QY 129 SVLQVWQPAINKGDMILAPLKLPLISGALSAMDVYG--TKMDRARGDLHYLSTSPDK 187  
DB 241 SEILVQSSAD-----FSSPSPITTTDIS--LSAAPLDQTSSESSFTTASAAAPVSSSTDYDG 293  
QY 188 LHDMAVSVKRNHPALGRQVDMKIAVQTFSAIIV-----VRI--- 225  
DB 294 SSASPVMSNMAGOIASSSSTIDNPTMSEFFLTSTEVDSDVSTVSALSAFLQSTTS 353  
QY 226 ----VLAPALASRPVQGVDFGVSTRGVLVANKGFD-----RMLS 263  
DB 354 NSFSTVSPSVFVPS--QSSSDVASSTANVSSS--FSDIPQSTSGSVVAQASALA 411  
QY 264 VQSRDOLRGAFVLGKKEPKALSEETDMLDAYKAIKSASVSGAIVNAG--KRWAGLPL 322  
DB 412 FQSTTEVYGAASSTMSLSTSLQSTT--LDSSSIASSSSDLTDVGVSTASIFL 469  
QY 323 DVATDGLKAVNSL-----VSATSLTKNGLALAGVAGVSKLQDMA 362  
DB 470 LSAERQASTSSFSVSPVAFVPSQSSDVASTASAPVSSFSFYTSIQAGGSSMTNPS 529  
QY 363 TKVI-----TDSATKAAVQSL--NLVGSVGFACMTTLAGLATDPAVKKAEFTQDKVST 416  
DB 530 SSTIVSSSTGSSSEEAATSTASATLSCSSSTY--VAQNLQSQP--PSTSLISE---SQ 581  
QY 417 ASSTTSVVAQDTYKLA--TVKMSGEA---ISSTGA-----SLASTVNLRHRSAPFA 465  
DB 582 ATSTAVLASSSVTTSPTTTAGAGATEASSLSTSTALETSQVTSQSTTALQTSFPASS 641  
QY 466 DIEGG-----ISAFSRSETPPL 484  
DB 642 STEGSETSQGFSTSVLVQMPSSISSEFSPOQTQM 680

RESULT 13  
HLVA\_SERNA

ID HLVA\_SERNA STANDARD; PRT; 1608 AA.  
AC P13320;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE Hemolysin precursor.  
GN SHLA.  
OS Serratia marcescens.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Serratia.  
OX NCBI\_TaxId=615;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.  
RC STRAIN=SN8;  
RX MEDLINE=88257037; PubMed=3290200;  
RA Poole K., Schiebel E., Braun V.;  
RT "Molecular characterization of the hemolysin determinant of Serratia marcescens";  
RL J. Bacteriol. 170:3177-3188(1988).  
CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.  
CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA REQUIRES SHLB FUNCTION.  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL; M22618; AAA50323.1; -.  
DR PIR; A28182; A28182.  
KW Hemolysin; Toxin; Outer membrane; Signal.  
FT SIGNAL 1 30  
FT CHAIN 31 1608 HEMOLYSIN.  
SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DA051 CRC64;

Query Match 5.0%; Score 119.5; DB 1; Length 1608;  
Best Local Similarity 20.5%; Pred. No. 17;  
Matches 120; Conservative 68; Mismatches 242; Indels 155; Gaps 21;

QY 8 QQPPQVAMESFRASDASLASVSVSTTSCRDQAITDYLKHHVFAAHREFSVIGSPDE 67  
DB 708 QYRAGLRERTHRSEKTRTENSASISGGSVK-LAEKDV---TFSGSKLVA---DK 758  
QY 68 RDAALANEOIDALVETRANRLYSEGETPATIET-----PAKAEKFDRLATASSAFEN 122  
DB 759 GDASVSGN-----KVSFLAADKRTASNTQRTIGGGTYTGGIDKLSSGVLEAGTEN 809  
QY 123 TFPAAASVQWQPAINKGDMILATPLKLPLISGALSAMDVYG--TKMDRARGDLHY-- 180



[illegible]



QY 459 -HRSAPADIE 468  
I I I:  
Db 394 LHNQGVRAID 404

Search completed: January 6, 2004, 09:45:44  
Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:36:48 ; Search time 39 Seconds  
(without alignments)  
3222.349 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407  
Sequence: 1 MHINQSAQQPPGVAMESFT.....EEGGISAFSRSETPQLRL 467

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacterioph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
-----				

1	1869.5	77.7	486	2	Q9JP34	Q9JP34 pseudomonas
2	1591.5	66.1	484	2	Q8RP03	Q8RP03 pseudomonas
3	460	19.1	518	16	Q8XCF0	Q8XCF0 ralsionia s
4	149.5	6.2	2155	16	Q8UPP9	Q8UPP9 agrobacteri
5	149	6.2	1009	16	Q8UBW1	Q8UBW1 agrobacteri
6	146	6.1	2089	16	Q92K98	Q92K98 rhizobium m
7	143.5	6.0	2055	2	Q85472	Q85472 abiotrophia
8	143	5.9	22152	4	Q8WK17	Q8WK17 homo saplen
9	142	5.9	4753	16	Q8PL13	Q8PL13 xanthomonas
10	141	5.9	536	17	Q9HPA2	Q9HPA2 halobacteri
11	140.5	5.8	642	16	Q8XSW6	Q8XSW6 ralsionia s
12	140.5	5.8	1983	16	Q931N9	Q931N9 staphylococ
13	140.5	5.8	2186	16	Q99TB0	Q99TB0 staphylococ
14	140	5.8	503	2	Q9FCQ5	Q9FCQ5 pseudomonas
15	140	5.8	2285	9	Q64046	Q64046 bacterioph
16	140	5.8	2285	16	Q31976	Q31976 bacillus su
17	139.5	5.8	491	2	Q93G71	Q93G71 campylobact
18	139.5	5.8	1795	16	Q91CJ9	Q91CJ9 staphylococ
19	138.5	5.8	491	2	Q8RTY4	Q8RTY4 campylobact
20	138.5	5.8	548	2	Q9RA74	Q9RA74 streptococ
21	138.5	5.8	3961	16	Q8P942	Q8P942 xanthomonas
22	138	5.7	1242	5	Q9VM05	Q9VM05 drosophila
23	138	5.7	1430	16	Q8PL12	Q8PL12 xanthomonas
24	138	5.7	1731	16	Q8UBW4	Q8UBW4 agrobacteri
25	138	5.7	2993	16	Q8NKS0	Q8NKS0 corynebacte
26	137.5	5.7	2283	2	Q8VQ99	Q8VQ99 staphylococ
27	137	5.7	1545	16	Q9RQ01	Q9RQ01 streptomyc
28	136	5.7	968	5	Q9VK10	Q9VK10 drosophila
29	135.5	5.6	1952	5	Q8MWM6	Q8MWM6 naegleria g
30	135.5	5.6	2186	16	Q8NM19	Q8NM19 staphylococ
31	135	5.6	973	16	Q8XQ04	Q8XQ04 escherichia
32	135	5.6	6077	12	Q8VAP1	Q8VAP1 white spot
33	135	5.6	6077	12	Q8QTB7	Q8QTB7 white spot
34	134.5	5.6	1307	16	Q8K792	Q8K792 streptococ
35	134.5	5.6	1713	3	Q8TGE1	Q8TGE1 saccharomyc
36	134.5	5.6	2016	5	Q9BIT0	Q9BIT0 plectococcus
37	134.5	5.6	2178	2	Q9KMR3	Q9KMR3 streptococ
38	134	5.6	2335	16	Q8ZDR6	Q8ZDR6 yersinia pe
39	134	5.6	2579	16	Q8DOR8	Q8DOR8 yersinia pe
40	133.5	5.5	1233	16	Q92SD7	Q92SD7 rhizobium m
41	133.5	5.5	1307	16	Q8PIPO	Q8PIPO streptococ
42	133.5	5.5	2124	16	Q98M03	Q98M03 rhizobium l
43	132.5	5.5	654	16	Q9A4U7	Q9A4U7 caulobacter
44	132	5.5	6077	12	Q91LB5	Q91LB5 white spot
45	131.5	5.5	553	2	Q912M2	Q912M2 rhodobacter

# ALIGNMENTS

RESULT 1			
ID	Q9JP34	PRELIMINARY;	PRT; 486 AA.
AC	Q9JP34:		
DT	01-OCT-2000	(TEMBLrel. 15, Created)	
DT	01-OCT-2000	(TEMBLrel. 15, Last sequence update)	
DT	01-DEC-2001	(TEMBLrel. 19, Last annotation update)	

DE	Hypothetical 50.7 kDa protein.	
OS	Pseudomonas syringae (pv. tomato).	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;	
OC	Pseudomonadaceae; Pseudomonas.	
OX	NCBI_TaxID=323;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=DC3000;	
RC	MEDLINE=98422476; PubMed=9748456;	
RA	Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,	
RA	Collmer A.;	
RT	"The Pseudomonas syringae pv. tomato HrpW protein has domains similar	
RT	to harpins and peptate lyases and can elicit the plant hypersensitive	
RT	response and bind to peptate.";	
RL	J. Bacteriol. 180:5211-5217(1998).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=DC3000;	
RX	MEDLINE=20243785; PubMed=10781092;	
RA	Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,	
RA	Petrici-OcMleja T., van Dijk K., Collmer A.;	
RT	"The Pseudomonas syringae Hrp pathogenicity island has a tripartite	
RT	mosaic structure composed of a cluster of type III secretion genes	
RT	bounded by exchangeable effector and conserved effector loci that	
RT	contribute to parasitic fitness and pathogenicity in plants.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=DC3000;	
RA	Ramos A.R., Rehm A.H., Collmer A.R.;	
RT	"Pseudomonas syringae pv. tomato DC3000 hrpL through hrpU";	
RT	Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.	
DR	EMBL; AF232004; AAF71504.1; -.	
KW	Hypothetical protein.	
SO	SEQUENCE 486 AA; 50709 MW; 91CFBC03397F8AD1 CRC64;	
Query Match		77.7%; Score 1869.5; DB 2; Length 486;
Best Local Similarity		78.4%; Pred. No. 1e-98;
Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;		
QY	1 MHINOSAQQPGVAKESFRASDASIASSVSSTTSCRDIAITDYLKHHVFAHRS 60	: : : : :       : : : : :       : : : : :       : : : : :
DB	1 MHINRQQPVTATDSFRASDASIASSVSSTTSCRDIAITDYLTQVFAHKL 60	: : : : :       : : : : :       : : : : :       : : : : :
QY	61 VTGSPDERDAALAHNEQIDALVETRAHRLVSEGETPATIAETFAKAEKFDRLATTASSAF 120	: :    : :    : :    : :    : :    : :    : :    : :    : :    : :
DB	61 PADSDAQAAVDVHNAQITAIETRASRLHFEGETPATIADTFEAKKIDRLATTSCAL 120	: :    : :    : :    : :    : :    : :    : :    : :    : :    : :
QY	121 ENTFFAAASVLYQKPAINKGWILATPLKLTPLISGALSGANDQGTQMDRARGDLY 180	: : : : :       : : : : :       : : : : :       : : : : :
DB	121 RATFFAAASLLQYKQPAINKGWILAPLKLPLPLISGALSGANDQGTQMDRARGDLY 180	: : : : :       : : : : :       : : : : :       : : : : :
QY	181 LSTSDKLDHAAVAASVKRHSPPALGQVWDKGIATQFSSALNVYVTLAPALASPPVQA 240	: : : : :       : : : : :       : : : : :       : : : : :
DB	181 LSAFSDRLDHAASVSRKHSPPSLARQVLDTCVAVQYTSNRNAVYVTLAPALASPPVQA 240	: : : : :       : : : : :       : : : : :       : : : : :
QY	241 VDEGVSTAGGLVANAAGFGDRMLSVQSRDQLRGAFVYLGKDKPEKKAALSEETDMLDAYXA 300	

Db 241 VDLGVSNAGGLAANAGFNRLLSVOSRDQRGAGLVYIGKDKPEKQSLSEEDIMLEAVKA 300  
QY 301 IKSASYSGAALNAGKRWAGLPDVAITDGLKAVRSLSVATSLTRNGIALAGVAGVSKLOK 360  
Db 301 IKSASYSGAALNAGKRWAGLPDVAITDGLKAVRSLSVATSLTRNGIALAGVAGVSKLOK 360  
QY 361 MATKNITDSATKAASQSLNLSVGVGFAGMTAGLATDPAVKRAESFIQDKVSTASST 420  
Db 361 MATKNITDPAITKAASQSLTNLAGSAVFAAGWTTAALTDPAAVKRAESFIQDTVKSTASST 420  
QY 421 TSVYADQTYKLAKTYKMGSEALISSTGASLSRSTNNLRHSAPADIEEGGISAFSRS-- 480  
Db 421 TSVYADQTYKLAKTYKMGSEALISSTGASLSRSTNNLRHSAPADIEEGGISAFSRS-- 479  
QY 481 PFQLRR 486  
Db 480 PFPRPR 485

RESULT 2  
Q8RP03 ID Q8RP03 PRELIMINARY; PRT; 484 AA.  
AC Q8RP03;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Type III effector HopTo1Pma.  
OS Pseudomonas syringae pv. maculicola.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=9511;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ES4326;  
RX MEDLINE=21862332; PubMed=11872842;  
RA Gutman D.S., Vinatzer B.A., Sarker S.F., Rana M.V., Kettler G.,  
Greenberg J.T.;  
RT "A Functional Screen for the Type III (Hrp) Secretome of the Plant  
Pathogen Pseudomonas syringae.";  
RL Science 295:1722-1726(2002).  
DR EMBL: AF458051; AAL84253.1; -  
SQ SEQUENCE 484 AA; 50520 MW; 4DB012DD5688163E CRC64;

Query Match 66.1%; Score 1591.5; DB 2; Length 484;  
Best Local Similarity 67.6%; Pred. No. 7,2e-83;  
Matches 330; Conservative 51; Mismatches 100; Indels 7; Gaps 3;

QY 1 MHINSAQQPPGVMESEFRTASDASLASSVSSTTSCRDLOAITYIKHHVFAHRFS 60  
Db 1 MYINSSISQSSIGTSFHSQ--SVASSSVAVSEAEQTKLMDITHTYLTDFVFAHRLP 58  
QY 61 VICSPERDALALNEQIDAVETFRANRLYSGETPATIAETFAKAEKEDRLATTASSAF 120  
Db 59 VNDIVDDVYHBAHVEIGKITSKALRLIDEGTALSIGETFAKAEKEDRRAASAGAL 118  
QY 121 ENTFFAASVLYQWQPAINKGDMLATPLPLPLISGALSAMDQVGTQMDRAHGDLY 180

Db 119 RATFFAASVLYQWQPAINKGDMLATPLPLPLISGALSAMDQVGTQMDRAHGDLY 178  
QY 181 LSTSPDKLDHAAVAVSRHSPALGRQVWDIGIANVTFSAIUVVRYTLAPALASPVQGA 240  
Db 179 LSTAPETCLDHAABASVRRHHPGVMBQAAADIGIANVQTYAARALRYTLAPALASPVQGA 238  
QY 241 VDFGVSTAAGLVANAGFGDRMLSVOSRDQLRGATVYLGKDKPEKQSLSEETIMLDAYKA 300  
Db 239 VDISSVSAAGSLVANAGFSEHMTVQARDHLRGCAEVLGIKDKQPKADLSEETIMLDAYKA 298  
QY 301 IKSASYSGAALNAGKRWAGLPDVAITDGLKAVRSLSVATSLTRNGIALAGVAGVSKLOK 360  
Db 299 IKSASYSGAALNAGKRWAGLPDVAITDGLKAVRSLSVATSLTRNGIALAGVAGVSKLOK 358  
QY 361 MATKNITDSATKAASQSLNLSVGVGFAGMTAGLATDPAVKRAESFIQDKVSTASST 420  
Db 359 MATKNITDPAITKAASQSLTNLAGSAVFAAGWTTAALTDPAAVKRAESFIQDTVKSTASST 418  
QY 421 TSVYADQTYKLAKTYKMGSEALISSTGASLSRSTNNLRHSAPADIEEGGISAFSRS-- 478  
Db 419 TSLADTYKLAKAGIDASAERIATTGASLRDFT--LRRTVREPDIEEGGVAAGTGSFV 475  
QY 479 ERPFQLRR 486  
Db 476 AVPEFQNR 483

RESULT 3  
Q8XQF0 ID Q8XQF0 PRELIMINARY; PRT; 518 AA.  
AC Q8XQF0;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Putative transmembrane protein.  
GN RSP1277 OR RS05322.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Ralstoniaceae; Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM11000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
Arlat M., Billault A., Brotier P., Camus J.C., Catolico L.,  
Chandler M., Cholme N., Claudel-Renard C., Cunne S., Demange N.,  
Gaspin C., Lavie M., Molson A., Robert C., Sautin W., Schlex T.,  
R. Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,  
Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
RL Nature 415:497-502(2002).  
DR EMBL: AL646083; CAD18428.1; -  
KW Plasmid; Complete proteome.  
SQ SEQUENCE 518 AA; 53581 MW; 82D486715F0F60EE CRC64;



Db 911 RTMALDMSFSGVEEKITDILDGRTA-----ALKSAVAGVEDRIAGALDSRTAALSG--- 961

QY 430 KLAKTVKDMGSAISSTGASLSTVNNLRHSRAPEADIEGGIS 473

Db 962 -IVSGAEERIAEALDSRTALDMTISGVEERIAEADARAASLS 1004

RESULT 5

Q8UBW1 PRELIMINARY; PRT; 1009 AA.

ID Q8UBW1

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Ice nucleation protein homolog.

GN INA OR ATU3977 OR AGR L.1758.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Rhizobium.

OX NCBI\_TaxID=176299;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21608550; PubMed=11743193;

RA Wood D.W., Sebuel J.C., Kaul R., Monks D.E., Kitajima J.P.,

RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,

RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. St.,

RA Chapman P., Clemmensen J., Decherage G., Gilliet W., Grant C.,

RA Kuyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,

RA Raymond C., Rouse G., Saenphlammach C., Wu Z., Romero P., Gordon D.,

RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krepan W., Perry M.,

RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,

RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,

RA Neeter E.W.;

RT "The genome of the natural genetic engineer Agrobacterium tumefaciens

RT C58.";

RL Science 294:2317-2323(2001).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21608551; PubMed=11743194;

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,

RA Querolio B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,

RA Houmlet K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,

RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,

RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,

RA Cielo C., Slater S.;

RT "Genome sequence of the plant pathogen and biotechnology agent

RT Agrobacterium tumefaciens C58.";

RL Science 294:2323-2328(2001).

DR EMBL; AE009328; AL44779.1; ALT\_INIT.

DR EMBL; AE008285; AAK89449.1; -.

KW Complete proteome.

SQ SEQUENCE 1009 AA; 100468 MW; 14D842830E8C65BE CRC64;

Query Match 6.2%; Score 149; DB 16; Length 1009;

Best Local Similarity 22.4%; Pred. No. 3.4;

Matches 127; Conservative 77; Mismatches 228; Indels 136; Gaps 25;

QY 7 AQP6GVANESFRFASDASLASSVRSVSTTSCROLOAI-TDYLKH---HVFAAHRSVYI 62

Db 118 ASQIDAMSTDOIKALNSSQVAGLSAQVATLSDELATFTDEIKSISANAIAGLSAAAI 177

QY 63 GSPDERPALANEOIDALVERFRANRLYSEGPATTAETFAAEKF-----DLATTA 116

Db 178 AGISTDMAAALTKSQIDAMSTDOIKALNSSQVAGLSAQVATLSDELATFTDEIKSISANAIAGLSAAAI 237

QY 117 SSAFENTPPPAASVLYQWQPAINKGWLATPLKPLTPLISGALSAMDQVGTQKMDRAG 176

Db 238 IALSKAQVSAALSTTQFAMSTDO-----IKALTSQVAGLSA--QVATLSSD---- 284

QY 177 DLYHLSSTPDKLHDAASVYKRHSFALGRQVDMGIA-----VQ 215

Db 285 ELALFSTDEIKAI GANNAVA-GLSAALAAALTTDMAALITQTQIAGLSSTQALNTANLA 343

QY 216 TFSALNVVFTVLAAPALASRPVQAGVDFGVSFAGGLVA-----NAGFDRLSVQSRDL 270

Db 344 TFSA-DEIKALISTKALA-----GLDYTKLSTQNALITKTOASLSSTQFAMSTDOI 395

QY 271 RG-CAFLVGMKDKPPKALSEE-----TDMLDAYKAIKSASYGAALNA----- 313

Db 396 KALTSSEQVAGLSAQVATLSDELATFTDEISAFSANAAYAGLSTPALLATITGNAATLT 455

QY 314 GRRVAGL---PLDYATDGLKAVRSIVSATSITRNGIAGAGVAGVSKIQKATKATITDGA 370

Db 456 KTOIAGLSSTQINLATSGISATFSADVKAISTR--ALAG-----LDYTKLSTGNAVA-AL 507

QY 371 KRAVSGSLNLVGSVGFAGMTT-----AGLATPBAVKAES---FTQDKVYS 415

Db 508 SDAQVSAALSTT-----QFAAMSTDOIKALTSSEQVAGLSAQVATLSDELATFTDEIKAI 562

QY 416 -----TASSTTSVADQTVKILANT-VKQSGEALISSTGASLSTVNNLRHSRAPE- 464

Db 563 IGANNAVAGLSAALALTTDNASALTQTQIAGKSSTQINL-----TSNULATFSADEI 616

QY 465 -----ADIEGGISAFSRSE 479

Db 617 KAITTKALGGLDVTTKLSTGINAALTKQ 644

RESULT 6

ID Q92K98 PRELIMINARY; PRT; 2089 AA.

AC Q92K98;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical transmembrane protein SMC00190.

GN R01816 OR SMC00190.

OS Bacterium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Sinorhizobium.

OX NCBI\_TaxID=382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Batloy-Hubler F., Guzy J., Bothe G., Ampe F., Batut J.,  
RA Boizard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godtje T., Goffeau A., Kahn D., Kiss E., Lelaure V., Maury D.,  
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Remperege U.,  
RA Renard C., Thebaud P., Vandenbol M., Welter S., Gallbert F.,  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL: AL591788; CAC6395.1; -.  
DR InterPro: IPR004089; Chtxaxis\_c1ned.  
DR InterPro: IPR006162; Ppanche\_attch.  
DR Pfam: PF00015; MCPsignal; 1.  
DR PROSITE: PS00012; PHOSPHOPANTHERINE; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 2089 AA; 222522 MW; 1D6334385A31ADMD CRC64;

Query Match 6.1%; Score 146; DB 16; Length 2089;  
Best Local Similarity 18.9%; Pred. No. 14;  
Matches 106; Conservative 84; Mismatches 208; Indels 162; Gaps 17;  
QY 3 INQSAQPPGVAMESFRITASDASSVSVSTTSCRDLOAITDYLGHNFAHRSFVI 62  
DB 188 LRSNARMAEYAM---RLAEPTNADRWMTVGGAVREVSNMGEIETIARATELEAL 244  
QY 63 GSPD-----ERDAALAHNEQIDALV-----E 83  
DB 245 VHSVEYALERSYENELRVRLVQELQLEALIGHS DRI RALVAGHTKXKODETASE 304  
QY 84 TRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAASVLYQWQPAINKDW 143  
DB 305 DIASRIAVSGEAFASLIDTRAA-----LTKSDHALEN----- 338  
QY 144 LAPLPLPLT-PLISG-----ALSAMDQVGTKMPBARAGDIHLYSTSP 185  
DB 339 LSTMLTTRIDALLSGLTAGVALNSNEFDARLDALSDVLTQREGDLSQFETRASTLIDANT 398  
QY 186 DKLDMAAVSVKRHSFALGRQVDMGIAVQFSALNVRTVLAPALASRPSVQGVADFGV 245  
DB 399 EKLNALINERARQNETLIARTDLNESLR-----IGQALISGGLDVIL 442  
QY 246 STAGGLVAMAGFGDRMLSVQSHD-----QLRGAF-----VLGKKEPEKA 286  
DB 443 SSJNSALDEKGSFROSLKSSADDAINDLDLKGFFEEKLQTTVQQLASAFDERFHEPAS 502  
QY 287 ALSEETDMLDAYKAIKSASYSGALNAGKRNAGLPIDVATDGLAKVSLVSATSLTKNGL 346  
DB 503 AFDKRAQSLDT-KIMESLHRINETVSGGSEALGGALDSSVD---KINSALSQSILT---L 555  
QY 347 ALAGGVAGVSKLQRPATKNITDSATKAAVSQLSNLVGS-----VGVFAGWTTA 394  
DB 556 ATALG-----ATQDFIEETIGRTSELISLIQAHNRIESVLSDTGSLGALTE 605  
QY 395 -----GLATDP---AVKAKSEFIQDKVSTASSSTTSYVADDTYKLANVYKMSGEA 442  
DB 606 AQGRINENGFGADALANALTTTSERSLITDGLDSRTSAFIEGLQSAHAHRIQTLTGSTDEI 665

QY 443 ISSTGASINSTVNNILRRRSA 462  
DB 666 TSAIMASQHRILNLTLSERTA 685  
RESULT 7  
ID 085472 PRELIMINARY; PRT; 2055 AA.  
AC 085472;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Extracellular matrix binding protein (fragment).  
GN EMB.  
OS Abiotrophia defectiva.  
OC Bacteria; Firmicutes; Lactobacillales; Aerococcaceae; Abiotrophia.  
OX NCBI\_TaxID=46125;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NVS--47;  
RX MEDLINE=99081722; PubMed=9864195;  
RA Manganello R., van de Rijn I.;  
RT "Characterization of emb, a gene encoding the major adhesin of  
RT Streptococcus defectivus";  
RL Infect. Immun. 67:50-56(1999).  
DR EMBL: AF067776; AAD03320.1; -.  
DR InterPro: IPR005877; Gpos\_YsIRK.  
DR Pfam: PF04650; YsIRK\_signal; 1.  
DR TIGRFAMs: TIGR01168; YsIRK\_signal; 1.  
FT NON\_TER 2055 2055  
SQ SEQUENCE 2055 AA; 215642 MW; 9699C11DDE93E2FD CRC64;

Query Match 6.0%; Score 143.5; DB 2; Length 2055;  
Best Local Similarity 22.2%; Pred. No. 20;  
Matches 123; Conservative 78; Mismatches 249; Indels 105; Gaps 23;  
QY 3 INQSAQPP-----PGVAMESFRITASDASLA--SSSVSVSTTSCRD--LQAI-----TDY 48  
DB 726 INEISQRPDLTFRKQAFMDQVETARDAVMAKVASAANNQAVTSARDQGLAVANNILPTPA 785  
QY 49 LGHNFAHRSFVIGSPDERDALAHN-----EQIDAL-----VETRANRLYSEGETPA 97  
DB 786 AKYPPALGH---VROADAKRQAIRDNANILTAEEQADLRQVDAQTTAAEAIAINQHTNA 842  
QY 98 TIAETFAKAEK-----DRLATTASSAFENTPFAASVLYQWQPAINKDML-----A 145  
DB 843 TLAKRADDGKALININPQPRSKPRAHQALEQV--AAKK-----ROAINNNQLTDEKXA 895  
QY 146 TPKEPLTPLISGA---LSGAMDQVGTKMPBARAGDIHLYSTSPDKLDMAAVSVKRHSFA 202  
DB 896 QAIQGVQDALANAKQVOAANDINGVNOAKTA--GTIAINNNINPQGRQKQAIAAIEALQ 954  
QY 203 LGRQVDMGIAVQFSALNVRTVLAPALASRPSVQ-----GAVDFGVSTAGGLVA 253  
DB 955 AKRLELEGRNDLITTEERNNALADLTAKQQAQKDAVNOQARNNTGAVAKKNGAGIQGINP 1014  
QY 254 NA-----GFGRMLSVQSHQDLRGAFVLGKDKPEKPAALSEETDML--DAYRA 300

RESULT 8	
Q8WX17	
ID Q8WX17	PRELIMINARY;
	PRT; 22152 AA

Query Match	Score	DB	Length
5.98	143	4	22152

• •  
• •  
      
• •  
• •  
• •  
  
• •  
  
• •  
  
• •  
  
• •  
            
  
• •  
• •  
      
  
         
         
         
  
• •

```

Db      4593  --IEGPTSSLTAGRKENSTLQEMNSTITESNIIISNVSGAITTEATIMEVPSFDATFIPT 4650
QY      362  ---ATK-----NITDSATKAASQL-----SNIVSGV-- 386

```

Db 4771 TSSPVSMSSVLTSVLKTAGKVDTSLEFVTSSPQSMNTLDDISVTSAAITDIE 4824

RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,

RA Farla J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Fernighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sene J.A.D., Silva C., de Souza R.F.,  
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezera R.I.D.,  
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
RA Seubald J.C., Kitejima J.P.,  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
RL Nature 417:459-463(2002).  
DR EMBL; AE01814; AAM3677.1; -.  
DR InterPro; IPR000688; Hypa.  
DR Pfam; PF01155; Hypa; 1.  
KW Complete proteome.  
SQ SEQUENCE 4753 AA; 471484 MW; ABC4B22EB09DBFE9 CRC64;

Query Match 5.9%; Score 142; DB 16; Length 4753;  
Best Local Similarity 22.6%; Pred. No. 79;  
Matches 115; Conservative 67; Mismatches 160; Indels 166; Gaps 22;  
QY 77 QDALVETRAVRLSEGETPATIAETFAKAEKFR---LATTSSAFENTPFAASVLOY 133  
DB 1017 QAOSLENTAAVLAGDDVALTAQAQSLNAQVQAADLOMTAASLDQSGASAKRTLRG 1076  
QY 134 MQPAI--NKGDMLATEPLKPL-----TPLSGALS--GMDVVG---TKMD 172  
DB 1077 TVAGTLNRKGLIARDAQLDAGTLRSSQGLSELSASVLSQGMQDQGVAAATFALQA 1136  
QY 173 RARGDLHLYSTSPKLDMAVSKRISRL--GRQVYDMGIAVTSALNV----- 223  
DB 1137 SAGGDLQ-----QAQSLKQASIALQAQDLTANG--SLQASASTLDLQSGRTAL 1183  
QY 224 -----RTVLPALPA-----SRPSVGAQVDFGVSTAGL-- 251  
DB 1184 NGAQSSAADATLNGATIAASAAVVLQSGASTTLAGDSIERGALDANADTLRSAGDITL 1243  
QY 252 --VANAQFGRMLSVQSRDQLRGAFVYLQKDKPEKQALSEETDMLDAYRAIK----- 302  
DB 1244 GGAQAQG---RDIALTATGILNNGAQVVAARDLGVQAA---SATNNSDATLGAKRDLRV 1298  
QY 303 -----SAS-----YSGAL-----NAKRNAGLPLDVATD 327  
DB 1299 AGGLNLGLSLHGERGLSLTSGSLQGRGVYSGDALSTASNGAFENAAQGLVSGKALSTAG 1358  
QY 328 GLKA---VRSLSVATSLT--KNGLALAGVAGVSKLQKQATKIITDSATKAAVSQSLNLYG 383  
DB 1359 SIYSNGQLGSVTGALTLTQNDIALQGVVSAATTLQATAGGDLQSGAGTLSA--QTVALLQA 1416  
QY 384 SVGFAGWTTAGLATDPAVKAESFIODRVKSTASTTSYVADQTVKLATVYKDSGEAI 443  
DB 1417 GRDLVAGGTL-----QASATLDVQAQRTLSL-----NQAQ 1447  
QY 444 SSTGASLSTVNLIRHRSAPENDEEG 471

DB 1448 AGTDASLHGN----RIATGRDAVLKSGG 1471

RESULT 10  
ID Q9HPA2 PRELIMINARY; PRT; 536 AA.  
AC Q9HPA2;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Htr17 transducer.  
GN HTR17 OR VNG1733G.  
OS Halobacterium sp. (strain NRC-1).  
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;  
OC Halobacteriaceae; Halobacterium.  
OX NCBI\_TaxID=64091;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahaffes G.G., Bergquist B., Pan M.,  
RA Shukla H.D., Laskey S.R., Balliga N.S., Thorsson V., Sprogon J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir J., Goo Y.A.,  
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenberger T.A., Beck R.F., Fohlshroder M., Spidlich J.L., Jung K.-H.,  
RA Alan M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ebnhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
RT "Genome sequence of Halobacterium species NRC-1";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
DR EMBL; AE005078; AAG19968.1; -.  
DR HSSP; P02942; 10U7.  
DR InterPro; IPR004089; Cmtaxis\_transd.  
DR InterPro; IPR003660; HAMP.  
DR InterPro; IPR004090; Me\_chemotaxis.  
DR Pfam; PF00015; MCPA1gnal; 1.  
DR PRINTS; PR00260; CHEMTNNDICR.  
DR SMART; SM00304; HAMP; 1.  
DR SMART; SM00283; MA; 1.  
DR PROSITE; PS01111; CHEMOTAXIS\_TRANSDC\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 536 AA; 56815 MW; 3B31D2FD8701EE04 CRC64;

Query Match 5.9%; Score 141; DB 17; Length 536;  
Best Local Similarity 23.2%; Pred. No. 3.9;  
Matches 119; Conservative 78; Mismatches 197; Indels 120; Gaps 23;  
QY 12 GVAMESFRTASDASLASSSVSVSTSCRDQAITYLKHNVFAARFVIGSPDERDA 71  
DB 63 GYVMDAST-----SLANVLGVVAISSTGTISYED-LKTSNAAR-----RRKDAEQAK 113  
QY 72 LAHQEQDALVETRAVRLSEGETPATIAETFAKAEKFRDLATTASSAFENTPFAASVYL 131  
DB 114 RADEQRAQQAQKAEQDADEAEETIAKQLQRAERDAQLSKKASEVSVWEKAAS-- 171  
QY 132 QYMPAINKGMMLATPLKPLPL--SGALSGMDVVG-----TQMDRAR----- 175  
DB 172 -----GD-----LTVANSSSDDAETVGAENENKMAAEEMAAARTREFAGD 214





GN MRP OR SAVI758.  
OS Staphylococcus aureus (strain M50 / ATCC 700699).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=158878;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iino T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hiraekawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus.";  
RL Lancet 357:1225-1240(2001).  
DR EMBL; AP003363; BAB57920.1; -.  
DR InterPro; IPR005613; AIP3.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF03915; AIP3; 1.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR TIGRFAMs; TIGR01168; YsIRK\_signal; 1.  
KW Complete proteome.  
SQ SEQUENCE 1983 AA; 215786 MW; 573CE6DFFC7FBF CRC64;  
  
Query Match 5.8%; Score 140.5; DB 16; Length 1983;  
Best Local Similarity 19.3%; Pred. No. 28;  
Matches 105; Conservative 94; Mismatches 203; Indels 143; Gaps 21;  
  
QY 20 TASDASL-----ASSSVRSVSTSCRDQAITDY LKHVFAAHRFSVIGSPDERDALAHN 75  
DB 930 TAANTNLIANNANADVQKTNALOGIQAITPATK-----VKTDAKNA----- 972  
QY 76 EOIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASAFENTPFAASVLYQWQ 135  
DB 973 --IDKSAETQNTITIRNND--ATIEEQAAQQLIDQAVATQAQIN-----AADDNQEVA 1023  
QY 136 PAINKGWLATPLKPLTPLISGALSQMDQVGTIKMDRAAGDLHYLSTSPDKLIDAVAS 195  
DB 1024 QAKDQQTQNIIVTIGPATQVKTDAKNA-----VNDKAREAITVINATFGATREKQEA 1075  
QY 196 VKRHSPLGRQVVDKGIADVFFSALNVRT-----VLAPALASRESVQGA----- 240  
DB 1076 INRWVTLNKRALTIDIGV-TSTTAAWMSIRDQAVNQIGAVQPHVTKKQATGVLNDLATK 1134  
QY 241 -----VDFGVSTAGGLVNAAGFGDRMLSVQSRDQLGSAF----- 275  
DB 1135 KOEINQNTNATTEKQVALINQVDOELATA--INNINQADTNAEYDQAQQLGTAKALNHIQ 1191  
QY 276 -----VLQMDKEPKAALSEETDMLDAVKAIKSASVGAALNAGKRNA----- 318  
DB 1192 PNIVKKPAAALQINQHNAAKLAELINATPATNTDEKNAALN--TLNQDROQALESIKNQNT 1249  
QY 319 GLPLD-----VATDGLKAVRSLSVATSLTNGALAGGYAGVSKLQMA--TKNITDSATK 372  
DB 1250 NAEVDOATVAENNIDAVQVDVVKQAARDKIT-----AEVAKRIEAVKQTPNATDEKQ 1304

QY 373 AAVSQSLNLVSGVGFAGMTTAGLATDPAVKKA-ESFTQDKVYSTASSTSYV-----ADQ 427  
DB 1305 AANNQINQI-----KQDALINQINQNTDQVDTTNGAANAIIDNVEAV 1348  
QY 428 TVK-----IAKTVDMSGEAISSTGASLSTNNLNRBSAPADIEEGGSAFSSSETP 481  
DB 1349 VIKRPAIADIEDAVKREKQDIDN-----SLQSDNEKEVASQALAKEKEKALALIDQQTN 1404  
QY 482 FQLAR 486  
DB 1405 SQWNG 1409  
  
RESULT 13  
Q997B0  
ID Q997B0 PRELIMINARY; PRT; 2186 AA.  
AC Q997B0;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein SA1577.  
DE SA1577.  
GN Staphylococcus aureus (strain N315).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=158879;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Iino T., Kanamori M.,  
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Uji Y.,  
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,  
RA Hiraekawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,  
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,  
RA Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus.";  
RL Lancet 357:1225-1240(2001).  
DR EMBL; AP003135; BAB42845.1; -.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR TIGRFAMs; TIGR01167; LPYTG\_anchor; 1.  
DR TIGRFAMs; TIGR01168; YsIRK\_signal; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 2186 AA; 238448 MW; 0CC84836F1868A4F CRC64;  
  
Query Match 5.8%; Score 140.5; DB 16; Length 2186;  
Best Local Similarity 19.3%; Pred. No. 32;  
Matches 105; Conservative 94; Mismatches 203; Indels 143; Gaps 21;  
  
QY 20 TASDASL-----ASSSVRSVSTSCRDQAITDY LKHVFAAHRFSVIGSPDERDALAHN 75  
DB 930 TAANTNLIANNANADVQKTNALOGIQAITPATK-----VKTDAKNA----- 972  
QY 76 EOIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASAFENTPFAASVLYQWQ 135

Db 973 --IDKSAETQHTNTIFNNDD--ATLEEQQAQAQQLLDQAVATAQONIN-----AADTNEVA 1023

Qy 136 PAINKGDLATPLKPLTPLISGALSGAMDQVGTMRDRAKGLHYLSTSPDKLHDMAVSV 195

Db 1024 QAMDQGTONTIVTGPATQVKTDAKNA-----VNDKAREALTNINATFGATREKGEA 1075

Qy 196 VKRHSPLGRQVDMGIAVQTFESALNVRT-----VLAVALASRPSVQGA----- 240

Db 1076 INRVNTLKNRALTIDGV--TSTTAMWNSIRDDAVNQIGAVQPHVTKQGTATGVINDLATK 1134

Qy 241 -----VDGVTAGGLVNAAGFQDRMLTSVQSRQJLAGGAF----- 275

Db 1135 KOEINQNTNATTEKQVALNVDDELATA---INNINQADTNAEVDAQQLGTKALINAIQ 1191

Qy 276 -----VLGMRKDEPKALSEETDMLDAVYKAISKASYSAGALNAKRMVA----- 318

Db 1192 PNIVKRPALAQINQNYNAKLAELINATPDATNDEKNAALN--TLNQDRQALESIKNQNT 1249

Qy 319 GLPLD---VATDGLKAVRSIVSATSITNGLALAGVAGVSKIQRNA--TKNITDSATK 372

Db 1250 NAEVDAQATVAENNIDAVQVDVVKQAKARDKIT-----AEVAKRIEAVKQTFNATDEKQ 1304

Qy 373 AAVSQSLNLVGSVGFAGWTAGLATDPVAKKA--ESFIDKVKKSTASTSVV-----ADQ 427

Db 1305 AAVNQINQL-----KQQAQINQINQNTDQVDTTNQAVNALINDVEAY 1348

Qy 428 TVK-----LAKTVKDMGSAISSTGASLRSTVNNLHRSAPAEADIEGGISAFSRSETP 481

Db 1349 VIKRPAKADIEXAKVKKQQAIDN-----SLDSTDMKVKVSQALAKKEXKALAIQQAQTN 1404

Qy 482 FQLR 486

Db 1405 SQVNO 1409

RESULT 14

Q9FCQ5 PRELIMINARY; PRT; 503 AA.

AC 09FCQ5;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE Flagellin (fragment).

GN FLIC.

OS Pseudomonas fragi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=296;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=NCIM8542;

RA Bellingham N.F., Morgan J.A.W., Saunders J.R., Winstanley C.;

RT "Flagellin gene sequence phylogeny in the genus Pseudomonas.";

RU Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

EMBL; AJ297934; CAC03723.1; -.

DR InterPro; IPR001492; FlagellinN.

DR InterPro; IPR001029; Flagellin\_C.

DR Pfam; PF00700; Flagellin\_C\_1.

DR Pfam; PF00669; Flagellin\_N\_1.

DR PRINTS; PR00207; FLAGELLIN.

DR ProDom; PD000316; Flagellin\_C\_2.

FT NON\_TER 1

FT NON\_TER 1

SQ SEQUENCE 503 AA; 51085 MW; AE42EEFB521BF7C1 CRC64;

Query Match 5.8%; Score 140; DB 2; Length 503;

Best Local Similarity 24.3%; Pred. No. 4.1;

Matches 99; Conservative 43; Mismatches 135; Indels 130; Gaps 19;

Qy 147 PLKPLTPLISGA-----LSGAMDQGTMRDRAKGLHYLSTSPDKLHDMAVSVK 197

Db 10 PVKGAVTIANNNSSIAQTAEAGN--QESTNITRLR-ELAQSANQDGKADDRASLQ 67

Qy 198 RHSPALG-----ROYVDMGIAVQTF----- 217

Db 68 EFTAVGELTRISSTTFGSRLLDGFSFQNSQFVGADANQGISFGKSDISATGLKSYG 127

Qy 218 --SALNVVRTVLAVALASRPSVQGAVDGVTAGGLVNAAGFQDRMLTSVQSRQJLAGGAF 275

Db 128 EASAAGVSTLSANV-----VGGANDAAFFKATGAAAFVAVDDQTLTINGTD----- 174

Qy 276 VLGMKDE-----PKAALSEETDMLDAVYKAISKASYSGA 309

Db 175 IMAKDSKIDAVNAEINKQTSKVTGASADAATGTLTISASDPFAVQSAASDAGFVA 234

Qy 310 ALNKGKPAQGLPLDADGLK---AVRSIV-SATSLTR-NGLALAGVAGVSKIQRNAK 364

Db 235 ATPA-KVLGRPEVEIQNGVKTITAGSLLENAAAATTAAATDSKTVGNASVKGDRLVTS 293

Qy 365 -----NITDSATKAASVQSLNLVGSVGFAGWTAGLATDPVAV-----KKAESFI 409

Db 294 EKQQAQINLADSTGTGSPGSLSKL---GLTAGSTQAKLINDTSVSNQGVKFKKDS-- 347

Qy 410 QDKVKSTASS-----TTSYADQTVKLAATVKMGSAISSTGASLR 451

Db 348 MDTIVSSINASTGVTASKNADMTLKLFST-KD-----ITTAGGSCR 388

RESULT 15

O6406 PRELIMINARY; PRT; 2285 AA.

AC O6406;

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Putative transglycosylase.

GN YOMI.

OS Bacteriophage SPBc2.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.

OX NCBI\_TaxID=66797;

RN (1)

RP SEQUENCE FROM N.A.

RA Lazarevic V., Duesterhoeft A., Solido B., Hilbert H., Manuel C.,

```

Oy      437 --DMSEALISSTGALSTVNNLR 458
          | | | | |
Db      839 GYDSQGNALTKNKLEKAEINTK 862

Search completed: January 6, 2004, 09:45:15
Job time : 43 secs

```

Search completed: January 6, 2004, 09:45:15  
Job time : 43 secs

Search completed: January 6, 2004, 09:45:15  
Job time : 43 secs